

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:34:29 : Search time 32.85 Seconds

(Without alignments) 141.819 Million cell updates/sec

Title: US-09-471-255-2_COPY_15_150

Perfect score: 711

Sequence: 1 IVSLICAVALNQHRSQENK.....DGKYYVYLKDAHADNVRFK 136

oring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	11.1	315	1 CALD_MOUSE	035887 mus musculus
2	76.5	10.8	622	1 PPID_HABIN	P44092 haemophilus
3	74.5	10.5	743	1 BGAL_THERT	P71989 thermotoga
4	74.5	10.5	836	1 UME6_YEAST	P39001 saccharomyc
5	74	10.4	710	1 PPCE_HUMAN	P48147 homo sapien
6	74	10.4	710	1 PPCE_PIG	P33657 sus scrofa
7	73.5	10.3	713	1 PRML_BRARE	P99233 brachydanio
8	73	10.3	113	1 SENA_APLCA	P92323 aplysia cal
9	73	10.3	417	1 Y943_MERJA	058353 methanococ
10	72	10.1	238	1 E6_GOSHI	001197 gossypium h
11	72	10.1	517	1 EAS_DROME	P34352 drosophila
12	72	10.1	1116	1 SLPH_BACBR	P38538 bacillus br
13	72	10.1	1592	1 GTF2_STRDO	P27470 streptococ
14	71.5	10.1	753	1 NECL_MOUSE	P21662 mus musculu
15	71.5	10.1	752	1 SLAP_ACEKI	P22258 acetogenium
16	71.5	10.1	903	1 ECIC_BOVIN	P54281 bos taurus
17	71	10.0	533	1 DHAX_YEAST	P22281 saccharomyc
18	70.5	9.9	546	1 STR_THENA	Q9X018 thermotoga
19	70.5	9.9	670	1 C21D_HUMAN	095447 homo sapien
20	70.5	9.9	700	1 CH60_PLAFG	P34940 plasmodium
21	70.5	9.9	899	1 FURS_DROME	P36016 drosophila
22	70.5	9.9	1101	1 FURS_DROME	P36016 drosophila
23	70	9.8	222	1 Y175_MERJA	Q57639 methanococ
24	70	9.8	948	1 SECA_SYNP7	Q55357 synechococ
25	69.5	9.8	235	1 SPEC_STRPY	P13380 streptococ
26	69.5	9.8	611	1 SNF1_CANGA	Q00372 candida gla
27	69.5	9.8	1376	1 RPOD_ARATH	P56764 arabidopsis
28	69.5	9.8	2366	1 TOXB_CLODI	P18177 clostridium
29	69	9.7	132	1 TDHG_SCHPO	Q92361 schistosach
30	69	9.7	430	1 OSTR_YEAST	P33767 saccharomyc
31	69	9.7	471	1 KP19_ARATH	Q39030 arabidopsis
32	69	9.7	660	1 P5207_MYCPN	P5207 mycoplasma
33	69	9.7	691	1 PP21_YEAST	P26570 saccharomyc

34	69	9.7	1597	1 GTF1_STRDO	P11001 streptococ
35	68.5	9.6	482	1 IFET5_HUMAN	Q13335 homo sapien
36	68.5	9.6	753	1 NECL_HUMAN	P29120 homo sapien
37	68	9.6	220	1 GSPB_ERWCH	Q01563 erwina chr
38	68	9.6	379	1 ACDS_CROAB	P52042 crostidum
39	68	9.6	352	1 HMD1_DICDI	P34135 dictyostell
40	68	9.6	719	1 YP62_CAEEL	Q09437 caenorhabdl
41	68	9.6	932	1 SECA_SYNP3	Q55709 synechocyst
42	68	9.6	1267	1 DHRI_YEAST	Q04217 saccharomyc
43	67.5	9.5	645	1 SDBE_BACSU	P16396 bacillus su
44	67.5	9.5	775	1 VP4_ROTIV	P11194 human rotav
45	67.5	9.5	798	1 ITB1_MOUSE	P09055 mus musculu

ALIGNMENTS

RESULT 1
ID CALD_MOUSE STANDARD; PRT; 315 AA.
AC 035887;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CALUMENIN PRECURSOR.
GN CALD;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RC STRAIN=ICR; TISSUE=Heart.
RX MEDLINE=97364750; PubMed=9218460.
RA Yabe D., Nakamura T., Kanazawa N., Tashiro K., Honjo T.;
RT "Calumenin, a Ca2+-binding protein retained in the endoplasmic
reticulum with a novel carboxyl-terminal sequence, hDEF.";
J. Biol. Chem. 272:18232-18239(1997).
RL J. Biol. Chem. 272:18232-18239(1997).
CC -1- FUNCTION: NOT KNOWN. BINDS CALCIUM.
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -1- SIMILARITY: BELONGS TO THE RETICULOCALBIN FAMILY.
CC -1- SIMILARITY: CONTAINS 6 EF-HAND CALCIUM-BINDING DOMAINS.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC EMBL, U01829; AAC53316.1; -
DR KMD; MG1:1097156; Calu.
DR InterPro; IPR002048; -
DR Pfam; PF00036; ehand; 2.
DR PROSITE; PS00018; EF_HAND_4;
KW Calcium-binding; Endoplasmic reticulum; Signal; Repeat.
FT SIGNAL 1 19
FT CHAIN 20 315
FT CA_BIND 81 92
FT CA_BIND 117 128
FT CA_BIND 164 175
FT CA_BIND 201 212
FT CA_BIND 242 253
FT CA_BIND 278 289
FT CARBOHYD 131 131
FT SITE 312 315
FT SEQUENCE 315 AA; 37063 MM; 742361814171E273 CRC64;
Query Match 11.18; Score 79; DB 1; Length 315;
Best Local Similarity 22.18; Pred. No. 4.3;
Matches 42; Conservative 32; Mismatches 46; Indels 70; Gaps 11;

```

QY 1 IVSLSLC-AVALANQ-----HRSOENKDNRRVSYDQSSQSKSENLTP 41
DB 7 LMCLSLCTAVALASPEKEDRVHHEPQLSDKVNDAQNFYDHAFL-GAEAKSFQQLT 65
QY 42 PDVYSGEGLQAEQIVYKID--QGYT-----SHGCHYHYNG-KVPIYD 83
DB 66 PESEKRLG---KIVSKIDDDKGFVTVDELKGMFAKRWIHEVYKQMGHDJNED 121
QY 84 ALPSEE-----LAKMDP-----NYOL-----KDAIVNEVGGYLIKY 116
DB 122 GLVSWSEYKNAATGYVLDPPDDGPFNYKOMVDERRRFMAKDGDGL-ATKEEFAFL 180
QY 117 DGKYYVYKID 126
DB 181 HPEEYDMKD 190

RESULT 2
PPID_HAEIN STANDARD; PRT: 622 AA.
ID PPID_HAEIN STANDARD; PRT: 622 AA.
AC P44092;
DE 01-NOV-1995 (Rel. 32, Created)
DE 15-JUL-1998 (Rel. 36, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
PEPTIDYL-PROLYL CIS-TRANS ISOMERASE D (EC 5.2.1.8) (PIPIASE D)
(ROTAMASE D).
PPID OR HI1004.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
NX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / K20 / ATCC 51907;
RX MEDLINE=95350630; Pubmed=7342800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierulff A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utechtack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghegan N.S.M.,
RA Greth C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT *whole-genome random sequencing and assembly of Haemophilus
RL Influenzae Rd. Science 269:496-512(1995).
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Balroch A.;
RL Unpublished observations (MAY-1998).
CC -1- FUNCTION: PIPASES ACCELERATE THE FOLDING OF PROTEINS. SEEMS TO BE
INVOLVED IN THE FOLDING OF OUTER MEMBRANE PROTEINS (BY
SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PPIC/PARVULIN FAMILY OF ROTAMASES.
STRONG, TO E.COLI PPID.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
INTRODUCED IN POSITION 43 TO PRODUCE THIS ORF.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC CC
DR EMBL: U37781; AAC22665.1; ALT_FRAME.
DR TIGR: H11004;

```

```

DR InterPro: IPR000297;
DR Pfam: PF00639; Rotamase; 1.
DR PROSITE: PS01096; PPIC_PPIASE_1; 1.
DR PROSITE: PS01098; PPIC_PPIASE_2; 1.
KW Isomerase; Rotamase; Transmembrane; Periplasmic.
FT DOMAIN 1 16 CTTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 37 POTENTIAL.
FT DOMAIN 38 62 PERIPLASMIC (POTENTIAL).
FT DOMAIN 270 356 PPIC-LIKE.
SQ SEQUENCE 622 AA; 69590 MW; EES90KA5381C4B4 CRC64;

```

Query Match 10.8%; Score 76.5; DB 1; Length 622;
 Best Local Similarity 21.0%; Pred. No. 16;
 Matches 37; Conservative 29; Mismatches 67; Indels 43; Gaps 6;

```

QY 1 IVSLSLC-AVALANQ-----KDNRRVSYDQSSQSKSENLTPDVYSGEGLQAEQI 56
DB 196 IATLSLADEMAKQSVSDDEIKTYEANKQSFVQEVQKVOYIDLADNISRNLQYVDVEI 235
QY 57 VIKITQ--GYTSHGCHYHYNGKVPYDALFSEELMDPNY---QLKDAIVNEVKG 110
DB 256 AQYQDNKAQFMTOHLAHIOFANEQ---DAKVAIEELQGANFADYAKAKSLDKISGNG 312
QY 111 GYI-----IKYDKRYVYL---KDAHADNVRXK 136
DB 313 GDLGWVNEMLPFAFEDAAALQVGQYSPINVDGNVHYLVQERRAKQSLQENKAO 368

```

```

RESULT 3
BGAL_THREET STANDARD; PRT: 743 AA.
ID BGAL_THREET STANDARD; PRT: 743 AA.
AC P77989;
DE 15-DEC-1998 (Rel. 37, Created)
DE 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACYASE).
GN LACZ OR LACA.
OS Thermomicrobacter ethanolicus (Clostridium thermohydrosulfuricum).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OX Thermomicrobacter group; Thermomicrobacter.
NX NCBI_Taxid=1757;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33223 / 39E;
RA Zverlov V.;
RT Submitted (FEB-1998) to the EMBL/GenBank/DBS databases.
RL -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC CC

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC CC
CC EMBL: Y08557; CAA69850.1;
DR InterPro: IPR001649;
DR Pfam: PF00703; Glyco_hydro_2; 1.
DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KW Hydrolase; Glycosidase.
FT ACT_SITE 388 388 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 453 453 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 743 AA; 85796 MW; FE011F51E51DFC CRC64;

```

Query Match 10.5%; Score 74.5; DB 1; Length 743;
 Best Local Similarity 27.8%; Pred. No. 31;
 Matches 25; Conservative 14; Mismatches 30; Indels 21; Gaps 4;

QY 44 QVSKKEGI-----GAEQIVIKINDGVYSHGHYHYNGKVPYDALFSEELM 92
 DB 198 EVINKEEVCSSVVTDIKEKQKEIVQIKNDNPLTWPHDPHYLVNSV---KLIAEVEIL 254
 QY 93 KDPYOLKADIVNEVGGYIIKRVKRYV 122
 DB 255 --DNVTKRG-----IKGVYERDDGKRYI 277
 RESULT 4
 UME6 YEAST STANDARD; PRT; 836 AA.
 ID UME6 YEAST STANDARD; PRT; 836 AA.
 P39001, 01-FEB-1995 (Rel. 31, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-NOV-1997 (Rel. 35, Last annotation update)
 DE TRANSCRIPTIONAL REGULATOR UME6 (NEGATIVE TRANSCRIPTIONAL REGULATOR OF
 TIME2).
 GN UME6 OR CAR80 OR CAR81 OR NIM2 OR YDR207C OR YB8142.04C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-S1278B;
 RA Smart W.C., Park H.-D., Cooper T.G.;
 RL Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RC SEQUENCE FROM N.A.
 RC MEDLINE=95011581; PubMed=7926768;
 RA Strich R., Surosky R.T., Steber C., Messenguy F., Dubois E.,
 RA Easton Esposito R.;
 RT "UME6 is a key regulator of nitrogen repression and meiotic
 development.";
 RT Genes Dev. 8:796-810(1994).
 RL [3]
 RC SEQUENCE FROM N.A.
 RC STRAIN-GNF88;
 RA Kumeno A.;
 RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RC SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Oliver K., Shore L., Harris D., Barrell B.G., Rajandream M.A.,
 RA Welsh S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RC CHARACTERIZATION.
 RC MEDLINE=92253412; PubMed=1579492;
 RA Park H.-D., Lucche R.M., Cooper T.G.;
 RT "The yeast UME6 gene product is required for transcriptional
 repression mediated by the CAR1 URS1 repressor binding site.";
 RT Nucleic Acids Res. 20:1909-1915(1992).
 RL [6]
 RC CHARACTERIZATION.
 RC MEDLINE=96100650; PubMed=8528081;
 RA Anderson S.F., Steber C.M., Easton Esposito R., Coleman J.E.;
 RT "UME6, a negative regulator of meiosis in Saccharomyces cerevisiae,
 RT contains a C-terminal Zn2/Cys6 binuclear cluster that binds the URS1
 RT DNA sequence in a zinc-dependent manner.";
 RT Protein Sci. 4:1832-1843(1995).
 RN [7]
 RC CHARACTERIZATION.
 RC MEDLINE=9618868; PubMed=8614637;
 RA Jackson J.C., Lopes J.M.;
 RT "The yeast UME6 gene is required for both negative and positive
 RT transcriptional regulation of phospholipid biosynthetic gene
 RT expression.";
 RT Nucleic Acids Res. 24:1322-1329(1996).
 CC -1- FUNCTION: BINDS TO THE URS1 SITE (5'-AGCGCGCGA-3') AND NEGATIVELY
 CC REGULATES THE EXPRESSION OF MANY GENES INCLUDING CAR1 (ARGINASE).

CC SEVERAL REQUIRED FOR SPOULATION, MATING TYPE SWITCHING, INOSITOL
 CC METABOLISM, AND OXIDATIVE CARBON METABOLISM. HAS BOTH A POSITIVE
 CC AND NEGATIVE ROLE IN REGULATING PHOSPHOLIPID BIOSYNTHESIS.
 CC -1- SUBUNIT: INTERACTS WITH IMEI.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
 CC CLUSTER DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J2186; AAA4471.1;
 CC EMBL: L24539; AAC14472.1;
 CC EMBL: D23663; BAA04890.1;
 CC EMBL: Z68194; CAA92346.1;
 CC HSSP: P12351; 1PYC.
 CC TRANSFAC: T01247;
 CC SGD: S0002615; UME6.
 CC InterPro: IPR001138;
 CC Pfam: PF00172; Zn_c1us; 1.
 CC PROSITE: PS00463; ZN2_CYS6_FUNGAL_1; 1.
 CC PROSITE: PS00463; ZN2_CYS6_FUNGAL_2; 1.
 CC Transcription regulation; Repressor; Activator; DNA-binding;
 CC Nuclear protein; Zinc; Metal-binding.
 CC FT DNA_BIND 771 798 ZN(2)-CYS(6), FUNGAL-TYPE.
 CC FT CONFLICT 101 101 V -> G (IN REF. 3).
 CC FT CONFLICT 363 363 I -> V (IN REF. 1).
 CC FT CONFLICT 443 443 N -> T (IN REF. 1).
 CC FT CONFLICT 465 465 G -> D (IN REF. 1).
 CC FT CONFLICT 465 465 G -> D (IN REF. 1).
 CC SQ SEQUENCE 836 AA; 91123 MW; 0DDA0A0B4A157182 CRC64;
 Query Match 10.5%; Score 74.5; DB 1; Length 836;
 Best Local Similarity 21.6%; Pred. No. 35;
 Matches 29; Conservative 23; Mismatches 43; Indels 39; Gaps 5;
 QY 11 LNCORSCQNDNNRVSIVDSGSSQSKSE-----NLTPDQVSO-REGIOAEQIVL 58
 DB 275 LSRNNSTNNNDNNSIOSDSRESNNNEIGYLRGTRKRGSPSSNOYVHNHDDQCAV 334
 QY 59 KIDQGVYSHGHYHYNGKVPYDALFSEELMKDNPOLKADIVNEVGGYIIKRVG 118
 DB 335 -----GVAPRNFYFNKDRREITDPNVXIDE-----NESKINISFVLS 371
 QY 119 KYTVYLKDAHADN 132
 DB 372 KY-----RDEAVSLN 381
 RESULT 5
 PCEL_HUMAN STANDARD; PRT; 710 AA.
 ID PCEL_HUMAN STANDARD; PRT; 710 AA.
 AC P48147;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE PROLYL ENDOPEPTIDASE (EC 3.4.21.26) (PROT-PROLINE CLEAVING ENZYME)
 DE (PE)
 GN PREP OR PEP.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RC MEDLINE=95047504; PubMed=7959018;
 RA Vanhoof G., Goossens F., Hendriks L., de Meester I., Hendriks D.,

RA Vriend G., van Broeckhoven C., Scharpe S.,
 RT "Cloning and sequence analysis of the gene encoding human lymphocyte
 RT prollyl endopeptidase."
 RT Gene 149:363-366(1994).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94375419; PubMed-8089089;
 RA Shiraawa Y., Osawa T., Hiraishima A.,
 RT "Molecular cloning and characterization of prollyl endopeptidase from
 RT human T cells."
 RT J. Biochem. 115:724-729(1994).
 CC -1- FUNCTION: CLEAVES PEPTIDE BONDS ON THE C-TERMINAL SIDE OF PROLYL
 CC RESIDUES WITHIN PEPTIDES THAT ARE UP TO APPROXIMATELY 30 AMINO
 CC ACIDS LONG.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PRO-|-XAA >> ALA-|-XAA IN
 CC OLIGOPEPTIDES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9A; ALSO KNOWN AS THE
 CC PROLYL OLIGOPEPTIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X74496; CA52605.1; -
 CC EMBL: D21102; BAA04661.1; -
 CC MEROPS: S09.001; -
 CC DR MIM: 600400; -
 CC DR InterPro: IPR001375; -
 CC DR InterPro: IPR002470; -
 CC DR InterPro: IPR002471; -
 CC DR Pfam: PF00326; Peptidase_S9; 1.
 CC DR PRINTS: PR00862; PROLIGOPASE.
 CC DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 CC KW Hydrolyase; Serine protease.
 CC FT ACT_SITE 554 554 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ACT_SITE 641 641 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ACT_SITE 680 680 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT CONFLICT 16 16 V -> I (IN REF. 2).
 CC FT CONFLICT 245 245 T -> C (IN REF. 2).
 CC FT CONFLICT 298 298 T -> A (IN REF. 2).
 CC FT CONFLICT 319 319 W -> R (IN REF. 2).
 CC FT CONFLICT 440 440 I -> L (IN REF. 2).
 CC FT CONFLICT 459 459 G -> S (IN REF. 2).
 CC FT CONFLICT 459 459 V -> I (IN REF. 2).
 CC FT CONFLICT 706 706
 CC SQ SEQUENCE 710 AA; 80763 MW; 241D0F2D761A8DD2 CRC64;

Query Match 10.4%; Score 74; DB 1; Length 710;
 Best Local Similarity 25.8%; Pred. No. 32;
 Matches 32; Conservative 14; Mismatches 46; Indels 32; Gaps 6;

DB 42 PDVVSOREGIAA-----EQTIVK-----ITDGYVTSR--GDHYHY----76
 DB 34 PDSEQTAFAVEAQNKIIVPLEOCPIRGKXKEMTELYDPKYSCHKKRRFFYFNTG 93
 DB 77 --NGKVPY--DAFSEELMKDPNYQLKADYINVEKGTITKDKYVYLLDAHADN 132
 DB 94 LQNRQVLYVODSLGEARVFLDPNILLSDGTVALR--GFAFSEDEYFAVGLSAGSDW 150
 QY 133 VRTR 136
 DB 151 VTRK 154
 RESULT 6
 PCRE_PIG ID PCRE_PIG STANDARD; PRT; 710 AA.
 AC P23687;

DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE PROLYL ENDOPEPTIDASE (EC 3.4.21.26) (POST-PROLINE CLEAVING ENZYME)
 DE (PR)
 GN PREP.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCBI_TaxID:9823;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ACTIVE SITE SER-554.
 RC TISSUE-Brain;
 RX MEDLINE-91152034; PubMed-1900195;
 RA Rennex D., Hemmings B.A., Hofsteenge J., Stone S.R.;
 RT "CDNA cloning of porcine brain prollyl endopeptidase and
 RT identification of the active-site seryl residue."
 RL Biochemistry 30:2195-2203(1991).
 [2]
 RP ACTIVE SITE HIS-680.
 RX MEDLINE-91291146; PubMed-2064618;
 RA Stone S.R., Rennex D., Wikstrom P., Shaw E., Hofsteenge J.;
 RT "Inactivation of prollyl endopeptidase by a peptidylchloromethane.
 RT Kinetics of inactivation and identification of sites of
 RT modification."
 RL Biochem. J. 276:837-840(1991).
 CC -1- FUNCTION: CLEAVES PEPTIDE BONDS ON THE C-TERMINAL SIDE OF PROLYL
 CC RESIDUES WITHIN PEPTIDES THAT ARE UP TO APPROXIMATELY 30 AMINO
 CC ACIDS LONG.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PRO-|-XAA >> ALA-|-XAA IN
 CC OLIGOPEPTIDES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: IN ALL TISSUES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9A; ALSO KNOWN AS THE
 CC PROLYL OLIGOPEPTIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M64227; AAA31110.1; -
 CC PIR: A37942; A37942.
 CC MEROPS: S09.001; -
 CC DR InterPro: IPR001375; -
 CC DR InterPro: IPR002470; -
 CC DR InterPro: IPR002471; -
 CC DR Pfam: PF00326; Peptidase_S9; 1.
 CC DR PRINTS: PR00862; PROLIGOPASE.
 CC DR PROSITE: PS00708; PRO_ENDOPEP_SER; 1.
 CC KW Hydrolyase; Serine protease.
 CC FT MOD_RES 1 1 BLOCKED.
 CC FT ACT_SITE 554 554 CHARGE RELAY SYSTEM.
 CC FT ACT_SITE 641 641 CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC FT ACT_SITE 680 680 CHARGE RELAY SYSTEM.
 CC FT VARIANT 29 29 A -> H.
 CC SQ SEQUENCE 710 AA; 80769 MW; 70286A86238D72C0 CRC64;

Query Match 10.4%; Score 74; DB 1; Length 710;
 Best Local Similarity 25.8%; Pred. No. 32;
 Matches 32; Conservative 14; Mismatches 46; Indels 32; Gaps 6;

DB 42 PDVVSOREGIAA-----EQTIVK-----ITDGYVTSR--GDHYHY----76
 DB 34 PDSEQTAFAVEAQNKIIVPLEOCPIRGKXKEMTELYDPKYSCHKKRRFFYFNTG 93
 DB 77 --NGKVPY--DAFSEELMKDPNYQLKADYINVEKGTITKDKYVYLLDAHADN 132
 DB 94 LQNRQVLYVODSLGEARVFLDPNILLSDGTVALR--GFAFSEDEYFAVGLSAGSDW 150

OY 133 VTRK 136
DB 151 VTRK 154

RESULT 7

PRML_BRARE STANDARD: PRT: 713 AA.

AC Q9W735;

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

PROMININ-LIKE PROTEIN (FRAGMENT)

Brachydanio rerio (Zebrafish) (Zebra danio).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Rasbora; Danio.

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

NCBI_Taxid=7955;

DB 283 LOANLSTVNSLRNALN 289

RESULT 8

SEN_APLCA STANDARD: PRT: 113 AA.

AC P29233;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

GN M0943.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 RX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RA MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kellera A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Nguyen A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
 RA Uitterlidge T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U67538; AAB98953.1;
 DR TIGR: M0943;
 DR Hypothetical protein; Transmembrane.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 148 168
 FT DOMAIN 399 415 PRO-RICH.
 SQ SEQUENCE 417 AA; 45355 MW; 6DF8D82E1AABD90 CRC64;
 Query Match 10.3%; Score 73; DB 1; Length 417;
 Best Local Similarity 22.9%; Pred. No. 21;
 Matches 33; Conservative 23; Mismatches 43; Indels 42; Gaps 6;
 QY 16 SQENKDNRRSYVDSSQSKSENLPDVOQKEIOAEIVIKITDGYVTSHGDR---72
 DB 52 NQENQNNNNENKESOSTNIGSEYKNEIKGNENHPLQ-----SQNENQNNNNNEE 105
 QY 73 -----YHTY-----NGKVPYDALFSEELMKDPYOLKADIVN 106
 DB 106 NENAMTNVGESEVYNNNEPANNYIEITPDGTRP-DKI-EEQWL-----YIKIVDPIV 158
 QY 107 EVKGGYIIIVDGRYYVYLKD 126
 DB 159 GGLAGIDIDYDGNITGLTD 178
 RESULT 10
 E6_GOSHI
 ID E6_GOSHI STANDARD; PRT; 238 AA.
 AC 001197;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUN-1998 (Rel. 36, Last annotation update)
 DE PROTEIN E6.
 GN E6.
 OS Gossypium hirsutum (Upland cotton).
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 CC Malvales; Malvaceae; Gossypium.
 CC NCBI_TaxID=3635;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN-CV. COKER 312; TISSUE-Fiber;

RX MEDLINE=92335179; PubMed=1631059;
 RA John M.E., Crow L.J.
 RT "Gene expression in cotton (Gossypium hirsutum L.) fiber: cloning of
 RT the mRNA.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:5769-5773(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COKER 312;
 RA MEDLINE=96178868; PubMed=8616253;
 RA John M.E.;
 RT "Structural characterization of genes corresponding to cotton fiber
 RT mRNA, E6: reduced E6 protein in transgenic plants by antisense
 RT gene.";
 RL Plant Mol. Biol. 30:297-306(1996).
 CC -1- TISSUE SPECIFICITY: IT IS PREDOMINANTLY EXPRESSED IN FIBER CELLS.
 CC -1- DEVELOPMENTAL STAGE: CONCENTRATION OF E6 IS HIGHEST DURING THE
 CC LATE PRIMARY CELL WALL AND EARLY CELL WALL SYNTHESIS STAGES.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-5, MET-10 OR MET-21 IS
 CC THE INITIATOR.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M92051; AAA33055.1;
 DR EMBL: M92051; AAA33056.1; ALT_INIT.
 DR EMBL: U30505; AAB03079.1;
 DR PIR: A46130; A46130.
 DR Cell wall.
 SQ SEQUENCE 238 AA; 28226 MW; 7162101A9A91AFB CRC64;
 Query Match 10.1%; Score 72; DB 1; Length 238;
 Best Local Similarity 26.4%; Pred. No. 13;
 Matches 28; Conservative 12; Mismatches 40; Indels 26; Gaps 5;
 QY 16 SQENKDNRRSYVDSSQSKSENLPDVOQKEIOAEIVIKITDGYVTSHGDRH 75
 DB 116 SSNNKDT-YYNNKNAYESKQOOL-----GEVI-----FEKGTSTENNNNN 158
 QY 76 YNKKVPYDALFSEELMKDPYOLKADIVNEVKGGIIVDGRYY 121
 DB 159 YNNGNNTNN--GEKQKMSDTRT-----LENGKYYIVKASENNY 195
 RESULT 11
 EAS_DROME
 ID EAS_DROME STANDARD; PRT; 517 AA.
 AC P54352; O9YX17;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ETHANOLAMINE KINASE (EC 2.7.1.82) (E6) (EASILY SHOCKED PROTEIN).
 GN EAS OR CG3525.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE-Embryo;
 RX MEDLINE=95007778; PubMed=7923374;
 RA Pavlidis P., Ramaswami M., Tanouye M.A.;
 RT "The drosophila easily shocked gene: a mutation in a phospholipid
 RT synthetic pathway causes seizure, neuronal failure, and paralysis.";
 RL Cell 79:23-33(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).

STRAIN-BERKELEY;
 RA MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Branton R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Bokoyva D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Buttle J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cwayle S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dev I., Dietz S.M.,
 RA Doonan K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paetle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smet H.,
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Styrcas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Wetsenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RA Science 287:2185-2195(2000).
 RA -1- CATALYTIC ACTIVITY: ATP + ETHANOLAMINE - ADP + O-
 RA PHOSPHOETHANOLAMINE.
 RA -1- PATHWAY: PHOSPHATIDYLETHANOLAMINE SYNTHESIS.
 RA -1- ALTERNATIVE PRODUCTS: 2 ISOPROPS; A LONG FORM (SHOWN HERE) AND A
 RA SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 RA -1- MISCELLANEOUS: MUTATIONS IN EAS CAUSES SEIZURE, NEURONAL FAILURE,
 RA AND PARALYSIS. THIS IS DUE TO AN EXCITABILITY DEFECT CAUSED BY
 RA ALTERED MEMBRANE PHOSPHOLIPID COMPOSITION.
 RA -1- SIMILARITY: BELONGS TO THE CHOLINE/ETHANOLAMINE KINASES FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L35603; AAC37209.1;
 DR EMBL; L35604; AAC37210.1;
 DR EMBL; AE003501; AAF48574.1;
 DR FLYBASE; FBgn000536; eas.
 DR InterPro; IPR002573;
 DR Pfam; PF01633; Choline_kinase; 1.
 DR Trnase; Trnase; Kinase; Alternative splicing.
 FT ACT SITE 374 374 BY SIMILARITY.
 FT VARSPLIT 130 151 MISSING (IN SHORT ISOFORM).
 FT SEQUENCE 517 AA; 59208 MW; BA35636B07D6786 CRC64;
 SQ

Query Match 10.1%; Score 72; DB 1; Length 517;
 Best Local Similarity 22.9%; Pred. No. 33;
 Matches 39; Conservative 16; Mismatches 43; Indels 72; Gaps 7;

3 SLICCAVALINCH-----RSQENKNNRVSTVDSGSSQSEN-----LTPDQ 44
 Db 27 SLISVROVNOQTSLSSQNOVONLNSNSNSPNSGSENNENBONSRIKAPED 86
 Qy 45 VSQKEGIAQEDIVIKITDQGYTSHGHYHYNGKVPDALFSE-----LAKD 94
 Db 87 KSRKEAI-----VPEVPFVEADVIQAKELKAYR 118
 Qy 95 PNYQAKDAD-----TVNEVKGIIKVDGKYVYLKDAHADNVK 136
 Db 119 PTWLSHVEKIRVPPQIEDRVSGP--KCDGD-----DDASFDTGTYNK 160
 RESULT 12
 SLPH_BACBR STANDARD; PRT; 1116 AA.
 ID SLPH_BACBR
 AC P38538;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE SURFACE LAYER PROTEIN PRECURSOR (HEXAGONAL WALL PROTEIN) (HWP).
 OS Bacillus brevis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/staphylococcus group; Brevibacillus.
 OX NCBI_TaxID-1393;
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 54-68.
 RC STRAIN-HPD31.
 RA EMBL; D90050; BAA14103.1;
 RA EMBL; A35129; A35129.
 RA EMBL; S. Tsudol A., Takagi H., Naruse Y., Yamagata H., Tsukagoshi N.,
 RA Ueda S.;
 RA "Conserved structures of cell wall protein genes among
 RA protein-producing Bacillus brevis strains."
 RA J. Bacteriol. 172:1312-1320(1990).
 RL -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 RL HEXAGONAL S-LAYER.
 CC -1- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D90050; BAA14103.1;
 DR EMBL; A35129; A35129.
 DR InterPro; IPR001119;
 DR Pfam; PF00395; SLH_2.
 DR ProSite; PS01072; SLH_DOMAIN; 2.
 KW Signal; Cell wall; S-layer; Repeat.
 FT SIGNAL 5 53
 FT CHAIN 54 1116 SURFACE LAYER PROTEIN.
 FT DOMAIN 57 120 SLH 1.
 FT DOMAIN 121 171 SLH 2.
 FT DOMAIN 172 231 SLH 3.
 FT SEQUENCE 1116 AA; 123397 MW; 86D583D7AC72546F CRC64;
 SQ

Query Match 10.1%; Score 72; DB 1; Length 1116;
 Best Local Similarity 28.1%; Pred. No. 84;
 Matches 34; Conservative 23; Mismatches 46; Indels 18; Gaps 8;
 Qy 22 NNRVSVGSSOSSSENVLPDQVSQKEGIAQEDIVIKITDQGYTSHGHYHYNGRY 80
 Db 872 NN--NHIDNTAS--KSALIMPEDEKQKGT--DKYVARVDE--VDGWTISLNTVADGKTQ 923
 Qy 81 PYDALFSEELMKDPNOLKADIVNEVKGIIYI--KYDGKTYVYLKDAHADNVK 135
 Db 924 KYTKASTAFI--DYYDLGLGIDGVD--GDYIWDISADIDGTFDVLVAVSSDDEIRT 979

OY 136 K 136
DB 980 Q 980

RESULT 13

GTFF2_STRDO STANDARD; PRT; 1592 AA.
ID GTFF2_STRDO
AC P27470;
DT 01-AUG-1992 (Rel. 23, Last Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
NCBI_TaxID=1317;
RX MEDLINE-91123227; PubMed-1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
Kagawa H.;
*peptide sequences for sucrose splitting and glucan binding within
Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
synthetase)*;
RA Bacteriol. 173:989-996(1991).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA
1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES
WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE), GTF-SI SYNTHESIZES BOTH
FORMS OF GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MOTANS.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90213; BAA14241.1;
PIR: A38175; A38175.
HSP: P00695; ZHEE.
InterPro: IPR002479;
Pfam: PF01473; CW-binding_1; 16.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-1.
FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).
FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1093 1592 6.5 X TANDEM REPEATS.
FT REPEAT 1158 1207 1.
FT REPEAT 1222 1272 2.
FT REPEAT 1287 1337 3.
FT REPEAT 1402 1451 4.
FT REPEAT 1514 1563 5.
FT REPEAT 1577 1592 6.
SO SEQUENCE 1592 AA; 176167 MW; BCAA66D079351ECF CRC64;
Query Match 10.18; Score 72; DB 1; Length 1592;

Best Local Similarity 23.28; Pred. No. 1.3e+02;
Matches 36; Conservative 18; Mismatches 45; Indels 56; Gaps 7;

OY 9 YALNQHSGEN-----KQNR---VSYYDGGSSGSKSEITFPDYQSOREGI 51
DB 1137 YYANDGRYENENGYQDFGDMRYFKDGNMNAVGLTYDGNV-----QTFDDKGV 1185
OY 52 QAEIVIKITPDQGYVTSQSHD-----THYNGK---VPYDALSEELIM 92
DB 1186 QAKRKII-VTHDGRVRFDDQNGNAVNTFTADTKGHRYIYLGKGVAVTGQVYQKOKLY 1244
OY 93 KDPYQKLDADIVNEVGGLIKYDGRYYVYLKDA 127
DB 1245 FEANGQ-----QKGDFTVSDGKLYFDYDVS 1271

RESULT 14

NECI_MOUSE STANDARD; PRT; 753 AA.
ID NECI_MOUSE
AC P21662; P22546;
DT 01-MAY-1991 (Rel. 18, Last Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROENDOCRINE CONVERTASE 1 PRECURSOR (EC 3.4.21.93) (PC1)
DE (PROHORMONE CONVERTASE 1) (PROPEPTIDE PROCESSING PROTEASE).
GN PCSK1 OR NECI OR NEC-1 OR ATP-1.
OS Mus musculus (Mouse), and Mus cookii.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090; 10098;
RX MEDLINE-91319778; PubMed-1862107;
RA Korner J., Chun J., Harter D., Axel R.;
*Isolation and functional expression of a mammalian prohormone
processing enzyme, murine prohormone convertase 1.*;
RT Proc. Natl. Acad. Sci. U.S.A. 88:6834-6838(1991).
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-MOUSE;
RX MEDLINE-91319778; PubMed-1862107;
RA Korner J., Chun J., Harter D., Axel R.;
*Isolation and functional expression of a mammalian prohormone
processing enzyme, murine prohormone convertase 1.*;
RT Proc. Natl. Acad. Sci. U.S.A. 88:6834-6838(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-MOUSE; STRAIN-LAF1;
RX MEDLINE-92041727; PubMed-1657897;
RA Nakayama K., Hosaka M., Hatuszawa K., Murekami K.;
*Cloning and primary sequence of a mouse candidate prohormone
convertase PC1 homologous to PC2, Furin, and Kex2: distinct
chromosomal localization and messenger RNA distribution in brain and
pituitary compared to PC2.*;
RT Mol. Endocrinol. 5:111-122(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-MOUSE;
RX MEDLINE-91203919; PubMed-2017186;
RA Seidah N.G., Marcinkiewicz M., Benjannet S., Gaspar L., Beaulieu G.,
Maret M.-G., Lazure C., Molloy M., Chretien M.;
*Cloning and primary sequence of a mouse candidate prohormone
convertase PC1 homologous to PC2, Furin, and Kex2: distinct
chromosomal localization and messenger RNA distribution in brain and
pituitary compared to PC2.*;
RT Mol. Endocrinol. 5:111-122(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES-MOUSE;
RX MEDLINE-9110525; PubMed-1988934;
RA Smeekens S.P., Avruch A.S., Lamendola J., Chan S.J., Steiner D.F.;
*Identification of a cDNA encoding a second putative prohormone
convertase related to PC2 in Atc20 cells and 13lets of Langerhans.*;
RT Proc. Natl. Acad. Sci. U.S.A. 88:340-344(1991).
RN [5]
RP SEQUENCE OF 214-478 FROM N.A.
RC SPECIES-MOUSE; TISSUE-Pituitary;
RX MEDLINE-91000356; PubMed-2169760;
RA Seidah N.G., Gaspar L., Mion P., Marcinkiewicz M., Molloy M.,
Chretien M.;
*cDNA sequence of two distinct pituitary proteins homologous to Kex2

RT and furin gene products: tissue-specific mRNAs encoding candidates
 RT for pro-hormone processing proteinases.";
 RL DNA Cell Biol. 9:415-424(1990).
 RN ERRATUM.
 RP SPECIES-MOUSE;
 RX MEDLINE-91090850; PubMed-2264933;
 RA Seidah N.G., Gaspar L., Mion P., Marcinkiewicz M., Milikay M.,
 Chretien M.; 9:789-798(1990)
 RL DNA Cell Biol. 9:789-798(1990)
 CC -1- FUNCTION: INVOLVED IN THE PROCESSING OF HORMONE AND OTHER PROTEIN
 CC PRECURSORS AT SITES COMPRISED OF PAIRS OF BASIC AMINO ACID
 CC RESIDUES. SUBSTRATES INCLUDE POMC, RENIN, ENKEPHALIN, DYNORPHIN,
 CC SOMATOSTATIN AND INSULIN.
 CC -1- CATALYTIC ACTIVITY: RELEASE OF PROTEIN HORMONES, NEUROPEPTIDES AND
 CC RENIN FROM THEIR PRECURSORS, GENERALLY BY CLEAVAGE OF -LIS-ARG-1-
 CC BONDS.
 CC -1- COFACTOR: CALCIUM DEPENDENT.
 CC -1- SUBCELLULAR LOCATION: LOCALIZED IN THE SECRETION GRANULES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY. HIGH SIMILARITY WITH OTHER FURIN-LIKE ENZYMES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 CC EMBL: M58507; AAA39896.1; -
 CC EMBL: X57088; CAA40368.1; ALT_SEQ.
 CC EMBL: M58589; AAA39894.1; -
 CC EMBL: M55668; AAA39375.1; ALT_SEQ.
 CC EMBL: M69196; AAA39732.1; -
 CC PIR: JX0171; KXMSC1.
 CC PIR: A37951; A37951.
 CC MEROPS: S08.072; -
 CC MD: MG1:97511; PcsK1.
 CC InterPro: IPR00209; -
 CC InterPro: IPR002884; -
 CC Pfam: PF01483; P. 1.
 CC Pfam: PF00082; Peptidase_S8; 1.
 CC PRINTS: PR00723; SUBTILASIN.
 CC PROSITE: PS00136; SUBTILASE_ASP; 1.
 CC PROSITE: PS00137; SUBTILASE_HIS; 1.
 CC PROSITE: PS00138; SUBTILASE_SER; 1.
 CC KMW Hydrolyase; Serine protease; Glycoprotein; Zymogen; Calcium; Signal.
 CC FT SIGNAL 1 27
 CC FT PROPEP 28 110
 CC FT CHAIN 111 753
 CC FT DOMAIN 122 410
 CC FT ACT_SITE 167 751
 CC FT ACT_SITE 167 167
 CC FT ACT_SITE 208 208
 CC FT ACT_SITE 382 382
 CC FT ACT_SITE 401 401
 CC FT CARBOHYD 645 645
 CC FT CARBOHYD 645 645
 CC FT CONFLICT 23 23
 CC FT SEQUENCE 753 AA; 84146 MW; 95878441BBBD9CD CRC64;
 SO
 Query Match 10.18; Score 71.5; DB 1; Length 753;
 Best Local Similarity 31.28; Pred. No. 58;
 Matches 20; Conservative 11; Mismatches 26; Indels 7; Gaps 2;
 QY 48 REGIOAGQYIKITDQGVYSHGDHYHYGKQVYDALFSEELMKRPYQLKADADIVNE 107
 DB 153 EKGITGQVYIVLDQGLMNNHTDIYANTDPEASTD--FNDNDHDPKRI-----DLTNE 205
 QY 108 VKGG 111
 DB 206 NKHG 209

RESULT 15
 SLAP_ACEKI STANDARD; PRT; 762 AA.
 AC P2258; 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE CELL SURFACE PROTEIN PRECURSOR (S-LAYER PROTEIN).
 OS Acetogenium kivi.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Thermotomabacter group; Thermotomabacter.
 OX NCBI_TaxID=2325;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-DSM 2030;
 RX MEDLINE-90036724; PubMed-2681162;
 RA Peters J., Peters M., Lottspeich F., Baumeister W.;
 RT "S-layer protein gene of Acetogenium kivi: cloning and expression in
 RL Escherichia coli and determination of the nucleotide sequence.";
 J. Bacteriol. 171:6307-6315(1989).
 RN [2]
 RP PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
 RC STRAIN-DSM 2030;
 RX MEDLINE-92281680; PubMed-1596358;
 RA Peters J., Rudolf S., Oeschkinet H., Mengele R., Sumper M.,
 RT Kellermann J., Lottspeich F., Baumeister W.;
 RT "Evidence for lysine-linked glycosaminoglycan in a bacterial
 RL surface protein.";
 J. Biol. Chem. Hoppe-Seyler 373:171-176(1992).
 RN [3]
 RP DOMAINS.
 RX MEDLINE-94156823; PubMed-8113161;
 RA Lupas A., Engelhardt H., Peters J., Santarius U., Volker S.,
 RA Baumeister W.;
 RT "Domain structure of the Acetogenium kivi surface layer revealed by
 RT electron crystallography and sequence analysis.";
 J. Bacteriol. 176:1224-1233(1994).
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
 CC OF PROTEINS WHICH CONT THE SURFACE OF BACTERIA.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- PTM: THE CARBOHYDRATE CONTENT OF THIS PROTEIN IS ABOUT 8% WHICH
 CC CORRESPOND TO ABOUT 40 TO 50 SUGAR MOLECULES PER MONOMER. O-LINKED
 CC GLYCANS CONSIST OF GLC, GALNAc AND GLCNAC.
 CC -1- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 CC EMBL: M31069; AAA21930.1; -
 CC PIR: A34355; A34355.
 CC InterPro: IPR001119; -
 CC Pfam: PF00395; SLH; 2.
 CC PROSITE: PS01072; SLH_DOMAIN; 2.
 CC KMW signal; Glycoprotein; Repeat; Cell wall; S-layer.
 CC FT SIGNAL 1 26
 CC FT CHAIN 27 762
 CC FT DOMAIN 30 94
 CC FT DOMAIN 95 145
 CC FT DOMAIN 146 204
 CC FT DOMAIN 473 479
 CC FT DOMAIN 625 630
 CC FT CARBOHYD 297 297
 CC FT CARBOHYD 516 516
 CC FT CARBOHYD 520 520
 CC FT CARBOHYD 632 632
 CC CELL SURFACE PROTEIN.
 CC SLH 1.
 CC SLH 2.
 CC SLH 3.
 CC SER/THR-RICH.
 CC SER/THR-RICH.
 CC O-LINKED (GLC...).
 CC O-LINKED (GLC...).
 CC O-LINKED (GLC...).
 CC O-LINKED (GLC...).

SO SEQUENCE 762 AA; 82785 MW; 34EC9C784DECA67E CRC64;

Query Match 10.1%; Score 71.5; DB 1; Length 762;

Best Local Similarity 23.8%; Pred. NO. 58; Matches 31; Conservative 19; Mismatches 41; Indels 39; Gaps 7;

QY 21 DNNRVSYVDGSQS-----SOKSENLP-----DOVSOKEGIOAEQIV 57
DB 279 DNDVVSFTGSDSVGTYVKKNDNNTAKVDNNAVLYNGYLTYSKYVKEGAE----- 333
QY 58 IKITDGGYVTSQDHYHYNGKVPY-DALFSEELMKDPNYOLKDADIVNEVKGGIIRV 116
DB 334 VTIINNNYLIIVNGS---YDNSTIYNDVQSGDKYLNRDSNYELKGVTVT---GAVSKV 386
QY 117 ---DKYIYY 123
DB 387 TDIKANDYIY 396

Search completed: September 26, 2001, 22:34:31
Job time: 235 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:32:23 ; Search time 46.65 seconds
(without alignments)
222.074 Million cell updates/sec

Title: US-09-471-255-2_COPY_15_150
Perfect score: 711
Sequence: 1 IVSLSLCAVALNHRSGENK.....DKRYYYLKDAAHADNVRK 136

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: pIR1.*
2: pIR2.*
3: pIR3.*
4: pIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	DB ID	Description
1	423	59.5	822	2	hypothetical 92.4K
2	85.5	12.0	721	2	virulence associat
3	83	11.7	814	2	DNA gyrase subunit
4	82	11.5	1946	2	lactococpin (EC 3.4
5	79.5	11.2	567	2	flagellar m-ring p
6	79.5	11.2	2573	2	hypothetical prote
7	78	11.0	262	2	triacylglycerol 11
8	78	11.0	438	2	spore cortex-lytic
9	78	11.0	438	2	probable flagellar
10	77.5	10.9	857	2	probable anthranil
11	77	10.8	1785	2	major merozoit su
12	76.5	10.8	594	2	hypothetical prote
13	76	10.7	875	2	ubiquitin-specific
14	75.5	10.6	312	2	leukocidin chain S
15	75.5	10.6	312	2	leucocidin chain S
16	75	10.5	312	2	hypothetical prote
17	75	10.5	1060	2	vitellinogen conve
18	74.5	10.5	388	2	GPR-binding protei
19	74.5	10.5	836	2	regulatory protein
20	74	10.4	710	1	prolyl oligopeptid
21	74	10.4	710	1	prolyl oligopeptid
22	74	10.4	710	2	prolyl oligopeptid
23	73.5	10.3	1028	2	hypothetical prote
24	73	10.3	113	2	sensorin A - Calif
25	73	10.3	282	2	protein p23 homolo
26	73	10.3	417	2	hypothetical prote
27	73	10.3	710	2	prolyl endopeptida
28	72.5	10.2	1712	2	hypothetical prote
29	72	10.1	205	2	ulcer-associated g

30	72	10.1	238	2	A46130	fiber protein E6 (
31	72	10.1	246	2	S65063	fiber protein E6 (
32	72	10.1	567	2	G64563	flagellar basal-bo
33	72	10.1	601	2	T32486	hypothetical prote
34	72	10.1	1116	2	A35129	surface layer prot
35	72	10.1	1592	2	A38175	glucosyltransferas
36	71.5	10.1	316	2	T19475	hypothetical prote
37	71.5	10.1	753	1	KXMS01	proteobactin convert
38	71.5	10.1	762	2	A34355	cell surface prote
39	71.5	10.1	1120	2	S67208	hypothetical prote
40	71.5	10.1	1714	2	E71609	Ser/Thr protein kl
41	71	10.0	478	2	A32555	major merozoit su
42	71	10.0	533	2	A39410	aldehyde dehydroge
43	71	10.0	3147	2	T18674	hypothetical prote
44	70.5	9.9	241	2	S65062	fiber protein E6 (
45	70.5	9.9	318	2	C64445	conserved hypothe

ALIGNMENTS

RESULT 1

T46758
hypothetical 92.4K protein - Streptococcus agalactiae

C:Species: Streptococcus agalactiae

C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000

C:Accession: T46758

R:Spellerberg, B., Rozdzinski, E., Martin, S., Weber-Heymann, J., Schitzler, N., J

Infect. Immun. 67, 871-878, 1999

A:Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attach

A:Reference number: 224091, MUID:99115568

A:Accession: T46758

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-822 <SPE>

A:References: EMBL:AF062533; NID:94249622; PIDN:RADJ3797.1; PID:94249624

A:Experimental source: strain R268

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 59.5% Score 423; DB 2; Length 822;
Best Local Similarity 55.1% Pred. No. 7.3e-30;
Matches 76; Conservative 26; Mismatches 34; Indels 2; Gaps 1;

OY 1 IVSLSLCAVALNHRSGENKNNRVSVDGSSQSKS--ENLTPDOVSKGIAQAEQIV 58
DB 15 ILAHHIGSYQKHHMGATKDNQATYDSSKGVKAKTKTKMDJSAEISAEQIV 74
OY 59 KITDGGVYTSQHDHYHYNGRVPYDALFSEELMKDPYQKADIVNEVGGYIKV 118
DB 75 KITDGGVYTSQHDHYHYNGRVPYDALFSEELMKDPYQKADIVNEVGGYIKV 134
OY 119 KRYYYLKDAAHADNVRK 136
DB 135 NYYYLKPGRSKRNIRK 152

RESULT 2

C82939
virulence associated protein, exoribonucleases U0057 [Imported] - Ureaplasma urealyticum

C:Species: Ureaplasma urealyticum

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: C82939

R:Glass, J.I., Lefkowitz, E.J., Glass, J.S., Helner, C.R., Chen, E.Y., Cassell, G.H.

submitted to GenBank, February 2000

A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a

A:Reference number: A82870

A:Accession: C82939

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-721 <GLA>

A:Cross-references: GB:AE002105; GB:AE222894; NID:96899003; PIDN:AAF30462.1; GSPDB:G

A:Experimental source: serovar 3; biovar 1

Db 235 ALBLOKRNENLEKINILAPIYGGKKNVARNAEEDFSQKSTKETEDPNVNR 294
 QY 46 -----SOKESIOAEQIVITITOGYTSIGDHYHNGVPPIDALFSEELMKDPYOL 99
 Db 295 SEQLLEKKEGAPKQYV-----GVPGVYVSNIGPVGGLDNKEPE-----KYEKSONTYVEV 347
 QY 100 KDADIVNEVKGYY-----IKVDGKYVYLKDAHA 130
 Db 348 -GRTISEIRGEFTGLVRLNAVAAYVVGKRYKIALEDGANA 384

RESULT 6

14
 Chemical protein PRB0460c - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: D71614
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 R:Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; M01D:99021743
 A:Accession: D71614
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2573 <GAR>
 A:Cross-references: GB:AE001366; GB:AE001362; NID:93845188; PIDN:AAC71881.1; PID:9384518
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PRB0460c

Query Match 11.2%; Score 79.5; DB 2; Length 2573;
 Best Local Similarity 20.8%; Pred. No. 1.2e+02;
 Matches 31; Conservative 25; Mismatches 38; Indels 55; Gaps 6;

QY 12 NOHRSQENKNNRVSIVDGSOSQKSF-----NLPPD-----QV 45
 Db 2421 NDNNDNNNDNNNSHTLAFQNRTOGETTETNNINNTDICEKGNKYTSNNNNINEM 2480
 46 SOKESIOAEQIVITITOGYTSIGDHYHNGVPPIDALFSEELMKDPYOL 99
 C:Species: Plasmodium falciparum
 C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: D71614
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 R:Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; M01D:99021743
 A:Accession: D71614
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2573 <GAR>
 A:Cross-references: GB:AE001366; GB:AE001362; NID:93845188; PIDN:AAC71881.1; PID:9384518
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PRB0460c

Query Match 11.0%; Score 78; DB 2; Length 438;
 Best Local Similarity 28.6%; Pred. No. 18;
 Matches 26; Conservative 10; Mismatches 27; Indels 28; Gaps 4;
 QY 58 IKITDQGVTSIGDHYHNGVPPIDALFSEELMKDPYOL 99
 Db 278 INVCCPMMTOWGSKRYLDEKGVYDILTSFGDLEKSAKRYGSPRSYGTTLTGY 337
 QY 103 -----DIVNEVKGYY-----IKVDGKY 120
 Db 338 GGPVRYVIOEOLNISRATPLIPKIAVDGKY 368

RESULT 9

Db1401
 Probable flagellar hook-associated protein Cj0548 [imported] - Campylobacter jejuni
 C:Species: Campylobacter jejuni
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
 C:Accession: D81401
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chli
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Be
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
 A:Reference number: A81250; M01D:20150912
 A:Accession: D81401
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-642 <PAR>
 A:Cross-references: GB:AL139075; GB:AL111168; NID:96967817; PIDN:CAW75184.1; PID:9696
 A:Experimental source: serotype O2, strain NCTC 11188
 C:Genetics:
 A:Gene: fliD; Cj0548

Query Match 11.08; Score 78; DB 2; Length 642;
Best Local Similarity 28.38; Pred. No. 29;
Matches 28; Conservative 16; Mismatches 39; Indels 16; Gaps 4;

QY 21 DNNRVSYDGSOSQSEKSENLPDQVSOKEGIAEQIVIKITDQGYVSH---GDHYHY 76
DB 84 DNPASLVNNGVALQSNINVTQAGKDYQSKGLA---NDGPFVNAQLNGTADLTFPS 140
QY 77 NGKVPYDALFSEELMKDPNTOLKD-ADIVNEVGGYII 114
DB 141 NGK-----EYTVYDKNTTYTDLADKINEASGEIY 171

RESULT 10

T04208
Probable anthranilate phosphoribosyltransferase (EC 2.4.2.18) TSC23.40 - Arabidopsis th
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 30-Apr-1999 #sequence, revision 30-Apr-1999 #text_change 11-Jun-1999
C/Accession: T04208
R/Author: M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, March 1999
A/Reference number: Z15361
A/Accession: T04208
A/Molecule type: DNA
A/Status: 1-857 <BEV>
A/Cross-references: EMBL:AL049500;
Experimental source: cultivar Columbia; BAC clone TSC23
C/Genetics:
A/Map position: 4
A/Introns: 150/3; 280/3; 391/2
A/Note: TSC23.40
C/Keywords: glycosyltransferase; pentosyltransferase

Query Match 10.98; Score 77.5; DB 2; Length 857;
Best Local Similarity 21.88; Pred. No. 47;
Matches 32; Conservative 25; Mismatches 39; Indels 51; Gaps 7;

QY 10 ALN-OHRSQEKDNNRSTYDGSOSQSEKSENLPDQVSOKEGIAEQIVIKITDQGYVTS 68
DB 3 AMNVEHRS---DKRHVY-MLPNSAOEHQHPQGNQSSSLAAEQ-----DN 46
QY 69 HGDRHYHYNGKVPYDALFSE-----ELMKDPNTOLKADIVNEVK 109
DB 47 HNEHHHHVPRKHOYDEKSEPARSKLYVHNSIASAPADALKEKTSPLHGGGVY---- 102
QY 110 GGYIIR-----VDGRYVYLR 125
DB 103 GGRVYHNDKRTATSTYDVERMYFLYVR 129

RESULT 11

T50182
merozolite surface antigen precursor - Plasmodium chabaudi chabaudi
C/Species: Plasmodium chabaudi chabaudi
C/Date: 30-Sep-1993 #sequence, revision 30-Sep-1993 #text_change 26-Aug-1999
C/Accession: A45546
R/Deleersnijder, W.; Hendrix, D.; Bendahman, N.; Hanegreets, J.; Briljs, L.; Hamers-Caster
Mol. Biochem. Parasitol. 43, 231-244, 1990
A/Title: Molecular cloning and sequence analysis of the gene encoding the major merozoit
A/Reference number: A45546; MID:91218805
A/Accession: A45546
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1785
A/Cross-references: GB:M34947; NID:9160597; PID:9160598
C/Keywords: major merozoite surface antigen
C/Keywords: surface antigen

Query Match 10.88; Score 77; DB 2; Length 1785;

Best Local Similarity 21.18; Pred. No. 1,3e+02;
Matches 23; Conservative 28; Mismatches 36; Indels 22; Gaps 5;

QY 21 DNNRVSYDGSOSQSEKSENLPDQVSOKEGIAEQIVIKITDQGYVSHGDH--YVN 77
DB 1104 DGNNTQIDPFKRSKEELNLTPEKVNQ-----LYLIDFAHKEISEHYNNRYK 1152
QY 78 GRVYDYLSEELMKDPNTOLKADIVNEVGGYIIR-----VDGRYV 122
DB 1153 YKLEERLYCKHEQIEAANKVKEISYK-----SRLAKRKRYINGTTYV 1197

RESULT 12

B64018
hypothetical protein HI1004 - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence, revision 18-Aug-1995 #text_change 29-Sep-1999
C/Accession: B64018
R/Reichmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Vente
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MID:95350630
A/Accession: B64018
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-594 <TIGR>
A/Cross-references: GB:U32781; GB:I42023; NID:91574028; PID:AMC22665.1; PID:91574034
C/Genetics:
A/Start codon: GTG
C/Superfamily: hypothetical protein HI1004

Query Match 10.88; Score 76.5; DB 2; Length 594;
Best Local Similarity 21.08; Pred. No. 36;
Matches 37; Conservative 29; Mismatches 67; Indels 43; Gaps 6;

QY 1 IVSISLCATVNLNHRSEN---KDNRVSYDGSOSQSEKSENLPDQVSOKEGIAEQI 56
DB 168 LATLSLADENAKQSVSDDEIKTYEANKQSPVOPEQVQYIDLSDNISRNLOVDVEI 227
QY 57 VIRITDQ--GYVSHGDRHYHYNGKVPYDALFSEELMKDPNT---QLKADIVNEVK 110
DB 228 AQYQDQKAPFQYOHIAHIOFANEQ---DAKVAEELOKANRADYAKASLDKISENG 284
QY 111 GYI-----IKYDGRYVYLR---KDAHADNRYTR 136
DB 285 GDLGWNENELPRAFEAAALQVQYSPINVDGNVHYLVQERRAQSLNRYKAQ 340

RESULT 13

T50182
ubiquitin-specific proteinase homolog [imported] - fission yeast (Schizosaccharomyces
C/Species: Schizosaccharomyces pombe
C/Date: 09-Jun-2000 #sequence, revision 09-Jun-2000 #text_change 09-Jun-2000
C/Accession: T50182
R/Padcock, K.; Churcher, C.M.; Wood, V.; Bartrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 2000
A/Reference number: Z25044
A/Accession: T50182
A/Status: preliminary; translated from GB/EMBL/DD8J
A/Molecule type: DNA
A/Residues: 1-875 <BAD>
A/Cross-references: EMBL:AL138854; PID:GAB72233.1; GSPDB:GN00066; SPDB:SPAC2363.08C
C/Genetics:
A/Experimental source: strain 97h(-); cosmid c2363
A/Gene: SPDB:SPAC2363.08C
A/Map position: 1

Query Match 10.78; Score 76; DB 2; Length 875;

Best Local Similarity 21.0%; Pred. No. 65;
 Matches 30; Conservative 27; Mismatches 46; Indels 40; Gaps 5;
 QY 2 VSLSLAVYALNORSEKNDNR-----VSYVDG-----SSQSKSE-----38
 DB 356 VNTDLSLSPNSHTSDNEDNEDYVSVSLVSEETDITYSKLSQSSPSOCHDSFL 415
 QY 39 --NLTPDQVQKEGIAEOIVIKITDQGYTSHG-----DHYHYNGKVPY 82
 DB 416 PANSFPLAASSTSLPSSSEILDSSSDKGOQVFGQHEVAGTNSFEDPNSHFNVSNSNHE 475
 QY 83 DALFSEELMKDPNYOLKADIV 105
 476 EASPKKEVL-KSPQFQRSLDIL 497

RESULT 14
 T00160
 leukocidin chain S precursor - staphylococcus aureus phage phi PVL
 N/Alternate names: protein 27
 C/Species: Staphylococcus aureus phage phi PVL
 C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
 C/Accession: T00160
 R/Kaneho, J.; Kilmura, T.; Kawakami, Y.; Tomita, T.; Kamio, Y.
 Biosci. Biotechnol. Biochem. 61, 1960-1962, 1997
 A/Title: Panton-Valentine leukocidin genes in a phage-like particle isolated from mitomy
 A/Reference number: Z14119; MOID:98067870
 A/Accession: T00160
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-312 <KAN>
 A/Cross-references: EMBL:AB009866; PIDN:BA31900.1
 C/Genetics:
 A/Gene: luks
 C/Function:
 A/Description: luks and luks cooperatively and strongly lyse rabbit erythrocytes besides
 C/Superfamily: leukocidin
 C/Keywords: hemolysis; toxin

Query Match 10.6%; Score 75.5; DB 2; Length 312;
 Best Local Similarity 23.0%; Pred. No. 20;
 Matches 26; Conservative 27; Mismatches 33; Indels 27; Gaps 6;
 QY 10 ALNQHRSQENKNNRVSYVDGSSQSKSENTPDQVSQKEGIG-----AEQIVIK 59
 DB 20 ATSFHESK--ADNNINENIGDAEYVRKTEPTSSDKNGVTYONIQDFVKKDKYKNDALILK 77
 QY 60 ITDQGYTSHGDHYHYNG-----KVPYDALFSEELMKDPNYOLKADIVN 106
 DB 78 M--QGFINSKTYTYNKNTPDHAKMRPQ--YNIGLKTNDPN-----VDLIN 121

RESULT 15
 S32211
 leukocidin chain S - Staphylococcus aureus
 C/Species: Staphylococcus aureus
 C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Mar-2000
 C/Accession: S32211
 R/Prevost, G.; Supersac, G.; Piemont, Y.
 submitted to the EMBL Data Library, March 1993
 A/Description: Staphylococcus aureus encodes two types of synergohymenotropic toxins: Pe
 A/Reference number: S32211
 A/Accession: S32211
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-312 <PRE>
 A/Cross-references: EMBL:X72700
 C/Genetics:
 A/Gene: luks
 C/Superfamily: leukocidin
 C/Keywords: toxin

Query Match 10.6%; Score 75.5; DB 2; Length 312;
 Best Local Similarity 23.0%; Pred. No. 20;
 Matches 26; Conservative 27; Mismatches 33; Indels 27; Gaps 6;
 QY 10 ALNQHRSQENKNNRVSYVDGSSQSKSENTPDQVSQKEGIG-----AEQIVIK 59
 DB 20 ATSFHESK--ADNNINENIGDAEYVRKTEPTSSDKNGVTYONIQDFVKKDKYKNDALILK 77
 QY 60 ITDQGYTSHGDHYHYNG-----KVPYDALFSEELMKDPNYOLKADIVN 106
 DB 78 M--QGFINSKTYTYNKNTPDHAKMRPQ--YNIGLKTNDPN-----VDLIN 121

Search completed: September 26, 2001, 22:32:25
 Job time: 309 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:30:36, Search time 32.85 Seconds
(without alignments)
1083.454 Million cell updates/sec

Title: US-09-471-255-2
Perfect score: 5406
Sequence: 1 MRFKRYIAAGSAVTVSLT.....IELRIPSGEVTKKNSDFIA 1039

ring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	227.5	4.2	1701	1 MSPI_PLAFM	P08569 plasmodium
2	226.5	4.2	1701	1 MSPI_PLAFM	P13819 plasmodium
3	205.5	3.8	1183	1 CNA_STAU	O53654 staphylococ
4	205.5	3.8	1726	1 MSPI_PLAFM	P04934 plasmodium
5	205.5	3.8	1726	1 MSPI_PLAFM	P04935 plasmodium
6	203.5	3.8	1637	1 MRSP_STAU	P80544 staphylococ
7	203.5	3.8	1637	1 IG44_HAEM	P45386 haemophilus
8	189	3.5	1682	1 MSPI_PLAFM	P19598 plasmodium
9	186.5	3.4	1570	1 P3K1_DICD1	P54673 dictyostell
10	185	3.4	2194	1 SC16_YEAST	P48415 saccharomyc
11	184.5	3.4	1466	1 SPA2_YEAST	P23201 saccharomyc
12	179.5	3.3	1636	1 BUD3_YEAST	P25558 saccharomyc
13	177	3.3	1702	1 IG42_HAEM	P45384 haemophilus
14	176.5	3.3	1139	1 HMW1_MYCGE	O49413 mycoplasma
15	176.5	3.3	1532	1 IGA_NEIGO	P09750 neisseria g
16	176.5	3.3	1658	1 YMG7_YEAST	O03661 saccharomyc
17	175.5	3.2	2748	1 NDM1_YEAST	Q00402 saccharomyc
18	174.5	3.2	1577	1 HLYA_PROMI	P16466 proteus mir
19	174.5	3.2	1807	1 VT42_XENIA	P18709 xenopus lae
20	174.5	3.2	2485	1 PTND_HUMAN	Q12923 homo sapien
21	173	3.2	1694	1 IG40_HAEM	P44969 haemophilus
22	173	3.2	1781	1 AKAC_HUMAN	Q02952 homo sapien
23	173	3.2	1790	1 USO1_YEAST	P25386 saccharomyc
24	172.5	3.2	1142	1 GIN4_YEAST	Q12263 saccharomyc
25	172	3.2	817	1 YG4A_YEAST	P46949 saccharomyc
26	172	3.2	2869	1 RBP1_PLAVB	O00788 plasmodium
27	171	3.2	710	1 L778_ARATH	O06738 arabidopsis
28	169.5	3.1	1167	1 SCPA_STRPY	P15926 streptococ
29	169	3.1	1744	1 TANA_XENIA	Q01350 xenopus lae
30	168.5	3.1	818	1 PKR2_YEAST	P47116 saccharomyc
31	168.5	3.1	1435	1 EBA1_PLAFC	P19214 plasmodium
32	168.5	3.1	1435	1 LTEL_YEAST	P07866 saccharomyc
33	168	3.1	1630	1 MSPI_PLAFM	P04932 plasmodium

34	168	3.1	3924	1 ANR2_HUMAN	Q01484 homo sapien
35	167	3.1	1639	1 MSPI_PLAFM	P04933 plasmodium
36	165.5	3.0	1256	1 ATL_STAU	P52081 staphylococ
37	164.5	3.0	1164	1 BAG_STRAG	P27951 streptococ
38	164.5	3.0	1189	1 YJH6_YEAST	P47035 saccharomyc
39	164	3.0	1222	1 S160_YEAST	P06105 saccharomyc
40	164	3.0	1679	1 YMG9_YEAST	Q04958 saccharomyc
41	163.5	3.0	1528	1 SPAA_STRDO	P21979 streptococ
42	163.5	3.0	1545	1 IG43_HAEM	P45385 haemophilus
43	163	3.0	1159	1 N124_SCHPO	O09904 schizosacch
44	163	3.0	1460	1 N159_YEAST	P40477 saccharomyc
45	162.5	3.0	857	1 NFM_CHICK	P16053 gallus gall

ALIGNMENTS

RESULT 1
MSPI_PLAFM STANDARD; PRT; 1701 AA.
ID MSPI_PLAFM
P08569:
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMNSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=70153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scalfe J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite
Plasmodium falciparum.";
RL J. Mol. Biol. 195;273-287(1987).
RN [2]
RP REVISIONS TO 1403; 1569 AND 1629.
RA Tanabe K.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scalfe J., Certa U.,
R Stuenkelberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
CC - SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
CC - PFM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SQUENCE OF 83 KDA, 42
KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC EMBL: X05624; CAA29112.1;
DR PIR: A26868; A26868;
DR PIR: B25120; B25120;
DR InterPro: IPR000561;
DR Pfam: PF00008; EGF_1;
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 20 1701 POTENTIAL.
FT CHAIN 1 19 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 4.28; Score 227.5; DB 1; Length 1701;
 Best Local Similarity 20.28; Pred. No. 0.005;
 Matches 250; Conservative 176; Mismatches 405; Indels 405; Gaps 66;

Query 1 MFSSKTYIAGSAVYISLCAVALNORSOEN---KDNRRVSVDSGSGSSENLTP 56
 500 MKNNNF---DKDYVDRKFFSARYYNEKORYNKRKSSNNSTYV---QKTKALSTLE 553
 57 DVSQKESIQAEQYVYKTTDGGYTSQGDHYHYNGRPYDA---LPSEEL----- 104
 554 D-YSLRKGI-----SERDFHYHYTLKTGLADIKLLEIKSEENKILE 596
 105 -----LMDPYOLKADIVN-EYKGGYII-KVDS---KYVYLLDAHADAVR----- 149
 597 KNFGGLHSANASLEVSIVYALQVOKVLLIKLIEDLRIELELNADLKOSIHVPNTYKP 656
 150 -----KDEINRKO-----EHVADNEKVNVA---VARSGRYTTNDG--- 185
 657 QNKPEPYLYLAKKEVDKKEFIKRYMDLKEQAVLSITQPLVASE---TTEDGHS 713
 186 -----YVNPADIIEDTGAAYVPHGCHYHYT-----PRSDLSASELAARAKHIA 230
 714 THTSOSGEETVEETEETEET-----VGHYTYTITLPRSEAPKEVKVENSIE 765
 231 GKNMOPSLASSTASDNNOGSVAKGSTSKPANK-----SENLOSLAKEL 275
 766 HK-----SNDNSOLAKTYVLLKLDLFLKSYICHKYLIVSSNSMOQKLELV 812
 276 YD-SPSAQRYSESDGLVFPKIIISRTPNGVALPHG-----DHYHFIYPSK----- 320
 813 YNLPEEKELEKS---CDPLDLFNQNNIPAMYSLYDSNNDOJOLFELVQEKMYTY 866
 321 LSALEEK-----IARVNPISGTSVST-NAPRNEVYSSLSISNPSLSLTSKELSA 373
 869 LHKLEENHKKLEBOKOITGSTSTSSPGNTVTYTAQSATHSNQSQSNASS---TNT 925
 374 SDGYTFNPKDIVEETATAYIRHGDHFIYIPKSNQI-GOPTLPY--NSLATPSPSLP 430
 926 QNG-----VAVSSGPAVVEESHPDPLTVSISMDLGIYSLNMGKRTVPPN-LTIST 977
 431 GTSHEKEDEGDFGDANRIIAEDESGFVMSHGDHNYFFKRD-----LLEEOIKAAQKH 494
 978 -TEMEKTYEN-----LAKNDTYF-----NDIKQFVANSNSVITGLFETQKNLANDE 1024
 485 LEVKYTS-----HN-----GLDSLSHEDYDEGNNAEKMD--LDKRIEKIAGIKMY 530
 1025 IRTKLDILOSFDLYNNYKULKLDRLFKRKKELOQDKMKIKLTELKQLESKLSNLNPH 1084
 531 GYKRE-SIVVNEKKNATII-----YPHGDHHDAPIDIEHNPVGI 567
 1085 NYLQNEVFNKKEKAEIATENTLENTLILKHYKGLVYTYNGE---SSPLTTLSEVSI 1141
 568 --GHSNLYELFPEEGVAKKEGKYTYGEE-----LTNVNVLKNSFPNNQFTLA 617
 1142 QTEBNYANLEKFRALSKIDKLDNDLHGKKKLSLSSGLHHLITELK-EVINKNYT-- 1198
 618 NGOKRVSFSPPELEKLGINMLVKL---ITPOKYLEKYSGVYFGEGVGNIANFELDOP 674

DB 1199 -----GNSPSENNKR--VNEALKSYENFPEAKVTTVVP-----POP 1234
 QY 675 YLPQTEFYRTI-----ASKDYPEVSYDQTFVPTSLAFKASQTIFFEPHAGDTYLRNP 729
 DB 1235 DVTSPSLSVKRYSGSGSKETQIPTSG-----SLTELOQVQOLONTDEEDSLVYLP 1288
 QY 730 QFAVPKGT-----ALVRYDFEPHGNVLENNYKVEIKLPKLNQGTTRAGN 779
 DB 1289 IFGSENDERELDQVWTEALSVYMDNLSG--FENEEDVYLLK-PL---AGVYRLSK 1341
 QY 780 KIP--VTFMANA-----YLD-----NOSTYVYEPILKENQTD 811
 DB 1342 QIERNITTFNLMNDIINSRLKRRYFLVLESDIOMFKHISSEYILE-DSFILNSEQ 1400
 QY 812 KPSLTPQFKRK-----AOENSKLDEY-----EERTSKYERKESLSEGNS 854
 DB 1401 KNTLAKSYKIKESVENDIKFAQGISYEEVLAQYDDELSIKYIKEREKPPSSPT 1460
 QY 855 TSNSTLEEVTPVDPOEKVAKFA-----ESYQKLENYLFN----- 891
 DB 1461 TPPE-----PAKDEQKREKFLPFLNIEFLYNNVKNKIDYILNKAINDCNVERKE 1515
 QY 892 -----DGTIELYPSGEVYIKRMADEFGAPQNGENRPSNGK-VSTGTV 936
 DB 1516 AHVYTKLSDKAIDKIDIF-----KNTNDEFAIKLINDDTKMDLKGILSTGLV 1567
 QY 937 ENOPT-----ENKRPDSLEAPNE--KPYKENS 963
 DB 1568 QNFPNTIISKLIEGKQMDLNTISOHCVKROCPENS 1603

RESULT 2

MSPL_PLAF 2
 ID MSPL_PLAF STANDARD; PRT; 1701 AA.
 AC P13819;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS) (PMSA).
 GN MSP-1.
 OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_Taxid=5837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88142999; PubMed=2449612;
 RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G., Brown G.V., Anders R.F., Kemp D.J.;
 RT Variation in the precursor to the major merozoite surface antigens of Plasmodium falciparum.
 RL Mol. Biochem. Parasitol. 27:291-302(1988).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POTENTIAL).
 CC -1- PTM: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
 CC CC
 CC EMBL: M19143; AAA29653.1;
 CC DR PIR: A54498; A54498.
 CC DR InterPro: IPR000561;
 CC DR Pfam: PF00008; EGF_1.
 CC KW Malaria; Merzoite; Polypeptide; Repeat; Signal; Glycoprotein;
 CC Transmembrane; GPI-anchor.
 CC SIGNAL 1
 CC 19 POTENTIAL.

FT CHAIN 20 1701 MEROZOTTE SURFACE PROTEIN 1.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SEQUENCE 1701 AA; 193719 MW; 39208/5E73D38552 CIRC64;

Query Match 4.28; Score 226.5; DB 1; Length 1701;
 Best Local Similarity 20.28; Pred. No. 0.0055;
 Matches 250; Conservative 175; Mismatches 406; Indels 405; Gaps 66;

QY 1 MKSEKTIAGASAVTISLCAVALNHRSGEN---KNNRVSVDGSSOSKSENITP 56
 DB 500 MKFNNNF---DKDVYDIFSAKYTYNEKORYNNKFSSNNSVYV---QRIKALSTLE 553
 QY 57 DVSQKREGIAEQIVITDQGYVTSQGDHYHYNGKVPYDA---LRFSEEL--- 104
 DB 554 D-YSLRNGI---SEKDNHYITLKTGLEADIKLITEKSEKNTLE 596
 QY 105 ---LAKDPYOLKADIVN-EYKGGYII-KVDG---KYVYLLDAAHADNVR--- 149
 DB 597 KNFGGLHSHANASLEVSQIVLQVQVLLIKKIDRLKIELFLNAQKDSIHVPIYKP 656
 QY 150 ---KDEINRKO---EHVKNERNVSNVA---VANSQGYTTNDG--- 185
 DB 657 QNKPEPYLLIYALKREVRLKEFIKVKDMLKEQAVLSITQPLVASE---TEDDGHS 713
 QY 186 ---YVNPADIEDTGNAYIYPHGGHYI---PKSDLSASELAAKAHIA 230
 DB 714 THYSQSGEETVETEETETET---VGHYTYTYITLPKESAPKVKYVENSIE 765
 QY 231 GKNMPOSLSYSTASDNTQSVAKGSTSKPANK---SENLSLKL 275
 DB 766 HK---SNDNSQALTKTYLKLDELFLKSYCHYIYVNSNMQKLELV 812
 QY 276 YD-SPSAQRYSQGLVDFPAKIISRPNGVALPHG---DAYHTPYSK--- 320
 DB 813 YNLPEEENELKS---CDPLDLLFNIONNIPAMYSITDSNNIDLOHLEFELYOKEMITY 868
 QY 321 LSALEEK---IARWPISGTGYSTV-TNAKNEVYSSIGLSNNSSLITREKELSA 373
 DB 869 LHKLEENHKKLEBOKQITGSTSTSPGWTYVTAQSAVSHSNQOONASS---TWT 925
 QY 374 SDGYIFNKDIVEETATAYIYRHGDHFHYIPKSNQI-GQPLFPN-NSLAPSPSLPINP 430
 DB 926 QNG---VAVSSGPRVVESSHDPPLVLSISNDLGIYVLSLNGNTKTPNP-LTIST 977
 QY 431 GTSHEKHEDGYGDANFIIADESGEYVMSHGDNNHFFKD---LREBQIKAAOKH 484
 DB 978 -TEMERFEEN---ILKNNDYF---NDDIKQFVKNSKVTYGLTETFOKALNDE 1024
 QY 485 LEEVTS---HN---GLDSSHEODYPGNAKEMD---LDKRIEKGIMQY 530
 DB 1025 IKKLDYQLQSLFDLYNKYKLLKDLRFNKKELGQDKMOKIKLITLKEQLESKLSLNP 1084
 QY 531 GVKRE-STVYNNKRNKAIIT---YPHGDHNADEIDEKRPYGI 567
 DB 1085 NTLQNFSTFPPNKKAEALAEENLENTKIKLHKGLYKTYNGS---SSPLKLTSEVSI 1141
 QY 568 ---GHSNHYELKPEEGVAKREKGNKYVIGEE---LTVNVNLLKNSFTFNQNTFLA 617
 DB 1142 QREDVYAMLEKPRALSKIDKLNLDLHGKKKLSFLSSGLHLLTELR-EVYIKNNYV-- 1198

QY 618 NGQKRVSPFPEPEKRLGIMLVKL---ITPDGKVLKYSQVYRGVGNINAFELDOP 674
 DB 1199 ---GNPSENNK--VNEALKSTENFLPEKATVTVTP-----PQP 1234
 QY 675 YLPQGTQKTYI---ASKDYPEVSYDGTFTVPSILAYKNASQTIFFPAGDTYLRVNP 729
 DB 1235 DVTPSPISLVRSVSGSGSTKEETQIPIISG-----SLTELOQVQVQANDEEDSLVLP 1288
 QY 730 QFVAPKQTD-----ALVRVDFEFGNAVLENNYKVEIKLPIPKINQGTTRAGN 779
 DB 1289 IFGESENDXYLDQVYVGEALSVTMDNLISG--FENEYDIYALK-PL---AGYRSILRK 1341
 QY 780 KIP---VTFANA-----YLD-----NOSTIYEPFALEKQTD 811
 DB 1342 QIEKNITTFMLNLDILNSRLKKRYFLDYLESIDMOKHJSSNEYLE-DSEFLINSEQ 1400
 QY 812 KPSILPQFKNK-----AOENSLDERV-----EPPTSKVEKESITGNS 854
 DB 1401 KNTLLKSYKIKESVENDIKFPAQGISIYEKVALKYDOLLSIKYIKEREKPPSSPT 1460
 QY 855 TSNSTLEVPYDPOEYKAFV-----ESYQKLENYLFNM----- 891
 DB 1461 TPFS---PAKDQKKESTFLPFLNIETLYNNLVNKNIDVILNKKAKINDGVENKE 1515
 QY 892 ---DGTIELYLPSEGVYIKNNMADFTGEADPOGNGENTPSENGK-VSTGTV 936
 DB 1516 AHVAKITSLDKALIDKIDLF-----KNTNDFEAIKILINDTFRKMDKILSTGLV 1567
 QY 937 ENQPT-----ENKRPDSELPAPNE--KPYKPPNS 963
 DB 1568 ONFENTITISKLEGGFODMINTISOHCYKRCPPNS 1603

RESULT 3
 ID CNA_STAU STANDARD; PRT; 1183 AA.
 AC 053654;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE COLLAGEN ADHESIN PRECURSOR.
 GN CNA.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID-1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SRRATIN-FDA 574;
 RX MEDLINE-92165839; PubMed-1311320;
 RA Patti J.M., Jonsson H., Guss B., Switalaski L.M., Wlberg K.,
 RT "Molecular characterization and expression of a gene encoding a
 Staphylococcus aureus collagen adhesin."
 RL J. Biol. Chem. 267:4766-4772(1992)
 RN [2]
 RP ERRATUM
 RA Patti J.M., Jonsson H., Guss B., Switalaski L.M., Wlberg K.,
 RL J. Biol. Chem. 269:11672-11672(1994).
 RN [3]
 RP COLLAGEN-BINDING DOMAIN.
 RC SRRATIN-FDA 574;
 RX MEDLINE-94032261; PubMed-8218209;
 RA Patti J.M., Boles J.O., Hoeck M.,
 RT "Identification and biochemical characterization of the ligand
 binding domain of the collagen adhesin from Staphylococcus aureus."
 RL Biochemistry 32:11428-11435(1993).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
 RX MEDLINE-97475225; PubMed-9334749;
 RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
 RA Moore D., Jin L., Schneider A., Delucas L.J., Hoeck M.,

RA Narayana S.V.L.;
 RT Structure of the collagen-binding domain from a Staphylococcus
 aureus adhesin.
 RL Net. Struct. Biol. 4:833-838(1997).
 CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
 COLLAGEN-CONTAINING SUBSTRATA.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, CELL WALL.
 CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 IN THE REGION OF THE MEMBRANE ANCHOR.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M81736; AAA20874.1;
 CC PDB: 1AMJ; 24-JUN-98.
 DR InterPro: IPR001899;
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
 DR Signal: Repeat; Transmembrane; Cell wall; 3D-structure.
 FT SIGNAL 1 29
 FT CHAIN 30 1183
 FT DOMAIN 30 1183
 FT TRANSMEM 1158 1177
 FT DOMAIN 1178 1183
 FT DOMAIN 151 318
 FT DOMAIN 533 1093
 FT DOMAIN 1093 1157
 FT DOMAIN 1151 1156
 FT REPEAT 533 719
 FT REPEAT 720 906
 FT REPEAT 907 1093
 FT SEQUENCE 1183 AA; 133066 MW; 6A1CC072E575D76 CRC64;
 Query Match 3.88; Score 205.5; DB 1; Length 1183;
 Best Local Similarity 19.78; Pred. No. 0.03; Mismatches 411; Indels 323; Gaps 50;
 Matches 214; Conservative 139; Mismatches 411; Indels 323; Gaps 50;
 QY 36 NNRVSYVDGSSQSKSENLTPDQVSOKEGIAEQIYIKITDQGYTSHGHYHYNGK 93
 DB 195 NNEKSYV-----SNDITIKQIQQGGQQLDSTLNTNT-----GHSNRYSGQS 238
 QY 94 --VPDALLESEELMKDPNTQLKADAT-----VNEVKGGYIIRVDG--KYVYLLDAA 142
 DB 239 AITDFEKAFPSKLTIVNTKNTIDVTIPQGYGSYNSFSINYKTKITNEQKREFFVNSQAW 298
 QY 143 HADNVRKDEINRKOEHYDNEVNSV-----ANARSGRTTN 183
 DB 299 YOEH--GKEEVNGSSEHNTYHINANAGISCTYKGLKLVLKQDKTRAPLANVFKLSKR 356
 QY 184 DGYVF--NPADITIEDTGNAVYIPHGCHYIYIKRSDLSASELAAKAHLAKGMQPSQL 239
 DB 357 DGSYVKNQKREIITDANGI-----ANKALPSSGYIILKEIAPRPYFDKDE---- 406
 QY 240 SYSTSTADNNTQSYAKSGTSKRPANKSENILSLKELDSPSAQRYSSDGLVDFDPAITIS 299
 DB 407 -YPTFMCDTDOGYF--TTIENAKAIKTKDY-----SAQR-----VNEGTOVKR 448
 QY 300 RTPGVAVIPRODHFHIFPKR--LSALEKIAKRVPISGSGSYSTNAKNEVYSSLSGL 357
 DB 449 PTIYFKLYKDDNONTTPVKAELKLEED-----GTTKYTWSTLPEND----- 491
 QY 358 SSNPSLTTSKELSSADGYIFNPKDIVEETATAYIVRHGDHFHYIKR--NOIGOPTLP- 415
 DB 492 -----KNKAKIKYLVKEVNAQGEPTTPE 514
 QY 416 -----NNSLATPSPSLDI-NGTSHEKHEDDYGGFDAANRIADDESGFVNSHGDHNTYFF 469

DB 515 GYTRKENGVLVNTNKEPIETTSISGERWMDKDNQDKR--PEKSVNLLANG----- 565
 QY 470 KKDLEBQIKAAQHLEEVKTSHGGLDLSLSEHEDYDGNAMKEMLDKITE-EKRIAGIMK 528
 DB 566 -----EKVKT-----LDVYSETNKT-----EFDLKRDKYDGKIEYTVT 600
 QY 529 QYVK-----RESIYVNR-----EKNAIYPPGDHHDAPIDERRPYGIGHSHSNYELFK 578
 DB 601 EDHKQDVTDDINGTTINKTTPGTSATVTKRNMDDNNQ--DGRPPEI-----KVLEY- 652
 QY 579 PEEGVA-----KKEGNTV-YTGEELIVNVILKSTNNQNT 615
 DB 653 -QDKAKGKTALINESNNMTHTWTGLDEKAKGQOVKTYVEELFVY-----KGYTHVNDND 707
 QY 616 LANGQKVSSEPELEKTKIGINLVKLLPDPGVLEKSVKVGEG-----VGNIAN 668
 DB 708 M--GNLIVTKNTPEETTSISGERWMDKDNQDKRPERKSVNLLADEKTKTLDVSETN 765
 QY 669 FELDQPILP-----GQTFKYTAS--RDYEVSTDGTFVPTSLAYKASQTIFFYPHAG 721
 DB 766 WKYEFKLPKRYDEKKEIETVTEDEHVDY--TTDINGT-----TINKKTPG 810
 QY 722 DTYLRV-----NQFVPPGTDALVYFDE-----FHGNATYENNYKV 759
 DB 811 ETSATVTKRNMDDNNQDKRPELEIKVELYDQKATGKTALINESNNMTHTWTGLDEKAKG 870
 QY 760 GEILPPIKLNQ--GTFRAGNKIPVTFMANVLDNQSYIYVPILEKENQTDK---- 812
 DB 871 QOVATYIEELKRYKGYTHHDNDGNLIVTKNTPEETTSISGERWMDKDNQDKRPERK 930
 QY 813 --PSILPOFRKRNK-----QENSKLDE--KVEEPTSEKVEKESLSENGST 855
 DB 931 VSVVLLANGKVKTLDTSETNMYEKKDLPKYDEGKIEYTVTEDEHVDKTYTDDINGTTI 990
 QY 856 SNGSL--EYVPTVD-----PVQRYAKFSESGMKLENLVFNQGTIELYLP 900
 DB 991 TNYTPGETSATVTKRNMDDNNQDKRPELEIKVELYDQKATGKTALINESNNMTHTWTG 1050
 QY 901 SGEYIKKNMADFT--GEAPDNGENKPSSENGKVSNGVYENQPTENKRPADSL--PEAPNKR- 956
 DB 1051 LDEKAKQOVKTYVDELTKVNGTTHVDNDGNLIVTKNTPEKPKPKPIYPEKPKDKTP 1110
 QY 957 PVKPENS 963
 DB 1111 PTPRDS 1117
 RESULT 4
 MSPL_PLAFC STANDARD; PRT; 1726 AA.
 AC P04934
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
 DE (PMSA) (P195).
 GN MSP-1
 OS Plasmodium falciparum (Isolate Camp / Malaysia).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID:5835;
 RN [1]
 RP SEQUENCE OF 1-1103 FROM N.A.
 RX MEDLINE-86205236; PubMed-3517809;
 RA Weber J.L., Leininger W.M., Lyon J.A.;
 RT Variation in the gene encoding a major merozoite surface antigen of
 RL the human malaria parasite Plasmodium falciparum.;
 Nucleic Acids Res. 14:3311-3323(1986).
 RN [2]
 RP SEQUENCE OF 1104-1726 FROM N.A.
 RX MEDLINE-88143999; PubMed-3278296;
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
 RT Merozoite surface protein sequence from the Camp strain of the human
 malaria parasite Plasmodium falciparum.;

RL Nucleic Acids Res. 16:1206-1206(1988).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL)
 CC -1- PPM: MEMOZITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEMOZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 EMBL: X03831; CA27446.1;
 DR A23386; SAZQM.
 DR InterPro: IPR000561;
 DR Pfam: PF00008; EGF, 1.
 KW Malaria; Membrane; Polyprotein; Repeat; signal; glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL: SURFACE PROTEIN 1.
 FT CHAIN 20 1726
 FT CARBOHYD 133 133 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 272 272 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 501 501 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 567 567 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 638 638 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 827 827 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 924 924 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 944 944 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 990 990 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC...) (POTENTIAL)
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC...) (POTENTIAL)
 FT SEQUENCE 1726 AA; 196197 MW; DDA045FA352BCF3 CRC64;
 Query Match 3.88; Score 205.5; DB 1; Length 1726;
 Best Local Similarity 20.08; Pred. No. 0.051;
 Matches 246; Conservative 172; Mismatches 414; Indels 397; Gaps 65;
 1 MKFSKTYIAGSANTVLSICAYALNDRSGEN---KDNRYGVDSOSSOSSENITP 56
 531 MKFNNE---DKDYVDKIFSAHYTYNEKORYNKFSSNSNVYV---QKLKALSTLE 584
 57 DOVSOKESIGAEQIVIKITDGYVTSQDHYHYNGKVPDA---LFSSEI----- 104
 585 D-YSLRKGI-----SEKDFNHITTLKTGLEADIKKLEIKSEKILE 627
 105 ---LMDPYOLKADIVN-EYKGYII-KVDS-KYVYVLDAAHADNRT----- 149
 628 KNFGLTSHANASLEVYDIYLVQKVLKIKIEDIKIELEKNAOLKDSIHVPNIKYP 687
 150 ---KDEINRKO-----EYVDNEKYNVNA---VASOGYITRDGYTF 188
 688 QNRPEYVYLYIKKEVDLKEPIPVKDKLKEQVYVLSIIPQVPAASE---TEEDG--- 741
 169 NPADIEDTGNAYIVPHGHHYIPKS-----DLASELAAKAHLAGNMOPSOISY 241
 742 ---GHSHTLSOSGEVETETETETETETETETETETETETETETETETETETET 784
 242 SSTA---SDNTQSVAGSTSKPANR-----SENIQSLKELYD-SPSA 281
 745 VENSLEHNSNDNSQALTTVYIAKIDELFTSYCHKIIVSNSMODKLEVYVLPFE 844
 282 QRTSSDGLVEDPAKIIIRTPNGVAIPRG-----DHYHEIPYSK-----LSALEEK 327
 845 ENELKS---CDPLDLNLNIONNIPAMTSLYDSNNNDLQHFELLYKREMYIYKLEKE 900
 328 ---IARMPISGTGTVST-NAPKEVYVSLGSLSNSNSLTSSELSASGCIYFN 380

Db 901 NHIKLLEOKOTIGTSSTSPGTYVNTAASATSHNSQOQSNASS---TNTONG----- 953
 Qy 381 KDIVETAAVAYVVRHGDHPIKRSQI-GQPLPN-NSLAPSPSLPDPNGTSHK 437
 Db 954 ---VAVSSGAVVEESDPLTVISINDGIYSLNIGKRTKVPNP-LTIST-TEMEKP 1008
 Qy 438 EEOGYGFARITAEDESGVMSGHNNHFFKDD-----LREOIKAKHLEKYS 491
 Db 1009 YEN-----ILKNDPTP---NDIKQFVNSKRYTGLTETQNNALNDEKIKEDT 1056
 Qy 492 ---HN-----GLDLSHEDDYPGNAMKD---LDKIEIKIAGIKQGVKRE-S 536
 Db 1057 LQISFDLYNNTYKIKLIDLEFNKKELGODKQKQIKKILKQLESKLSNPNHVLQNF 1116
 Qy 537 IVNKKENALYI-----YPHGDHHRADPIDERKPVGI-GSHS 572
 Db 1117 VFFRKKEALAELENTLENTKILKRYKGVKYNE---SSPLKTLSEVSIQTEDNYA 1173
 Qy 573 NYELKPEGVAKKEGKVTYGE-----LTVNVLKNSFPNNQNTLNGQKRV 624
 Db 1174 NLEKFRVLSIDGKLNLDNLHGKRLKSLFSGHLHLLTEIK-EYIKRNTY----- 1223
 Qy 625 FSPPELEKLGIMVYL---ITPDGYLEVSGKVFSGVGNIANFELDQYLPQOTF 681
 Db 1224 GNSPSENNKK---VNEALKSTENLPKAVTVVTP-----PQDPTSP 1266
 Qy 682 KYTI-----ASKDYEVSYDGFVPTSLAYKASQTFYFPHAGDYLYRVQFAVPK 736
 Db 1267 SVRVSSGSGSTKEHTQIPTSG-----SLTELOVVOLOANYDESDSLVLPFGESD 1320
 Qy 737 TD-----ALRVDFEFGNAVLENNYKGEIKLPKLNQGTTRRANKIP---V 783
 Db 1321 NDELDQVTVGEALSVTMNLLSG---FENEVDYTLK-PL---AGYRSIKQIKENIF 1373
 Qy 784 TPANA-----YLD-----NSTYIVYVPILEKNOTKPSILPQ 818
 Db 1374 TFINLNDIINSRLKRRKRYLDVLESDLQPKHISSEYITE-DSFILNSEQNTYLLKS 1432
 Qy 819 FKRNR-----AOENSKLDEKV-----EPTSEKEREKLSGTGNSNSTLE 861
 Db 1433 KYIKESVENDIKRAOGISITYEVNLAKYKDDLESIKYIKEREKPPSPPTTPPS--- 1489
 Qy 862 EVPTVDVQEVKAFNA-----ESYGRKLENTVLFM----- 891
 Db 1490 ---PAKTQOKRESKFLPFLNIEITLYNNLVKIKIDDYILNKAINDCQVEDEAHVITK 1547
 Qy 892 ---DGTIELPISGEVIAKMAADFTGAPGNGENBPSENGK-VSTGYVNOPT-- 941
 Db 1548 LSDKALIDKIDLF-----KNHNDFEALIKKLINDYKMDLIGLSTGLVONFPRTI 1599
 Qy 942 ---ENKPADSLPEAPNE-KYKPPNS 963
 Db 1600 ISKLIBEFQDMINISQHCYKOCPPNS 1628
 RESULT 5
 MSPL_PLAAP STANDARD; PRT; 1726 AA.
 AC P50495;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEMOZITE SURFACE PROTEIN 1 PRECURSOR (MEMOZITE SURFACE ANTIGENS)
 DE (PMSA) (GP195).
 GN MSP-1.
 OS Plasmodium falciparum (Isolate Palo Alto / Uganda).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=51270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99005525; PubMed=3049134;
 RA Chang S.P., Kramer R.J., Yamaga K.M., Kato A., Case S.E.,
 Siddiqui W.A.;

RA Savolainen K., Kussela P., Paulin L., Korhonen T.K.:
 "A large repeat-rich surface protein of methicillin resistant
 Staphylococcus aureus." Submitted (Dec-1998) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP PARTIAL SEQUENCE.
 RC STRAIN-ISOLATE 1061;
 RX MEDLINE-96670743; PubMed-8665912;
 RA Hilden P., Savolainen K., Tynnelae J., Vuento M., Kussela P.:
 "Purification and characterization of a plasmin-sensitive surface
 protein of Staphylococcus aureus." Eur. J. Biochem. 236:904-910(1996).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF15379; AAD09131.1;
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
 KW Cell wall; Transmembrane; Antibiotic resistance; Glycoprotein; Repeat;
 KW Signal.
 FT SIGNAL 1 48 POTENTIAL.
 FT CHAIN 49 1637 METHICILLIN-RESISTANT SURFACE PROTEIN.
 FT DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-(SAG).
 FT DOMAIN 1598 1603 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 SO SEQUENCE 1637 AA; 174573 MW; 75BE9ADB46BD309 CRC64;
 Query Match 3.8%; Score 203.5; DB 1; Length 1637;
 Best Local Similarity 19.0%; Pred. No. 0.058;
 Matches 228; Conservative 144; Mismatches 462; Indels 365; Gaps 50;

QY 2 KRSKRTIANGSA-VIYSLSLC-----AYALNQHSGENKNNRKYVDGS-----QS 47
 DB 18 KKSIRFTVGTSLVIGALILFGVANDQAELENNNTOKODSSDSKYGAVQTIEGSS 77
 QY 48 SOKSEMLPDDOYSOKGIGIAEOIVIKITDOGYVTSBGDHYHYNGRYVPDALFSEELMK 107
 DB 78 ANSNESDIPEQVD-----VTKDTE-----QASTEEKANTT 108
 QY 108 DPNYQDKADIYENVGYYLIKVDKYYVLLKDAHADNVRTKDEINROKOEHVKNRY 167
 DB 109 EQASTEKADVTTEQATTEAPRAEETGDKYTEAPRAEET---DKATTEAPRAEETDKA 165
 QY 168 NSNVAVARSGKRTTNDGYVFPADIIEDTGNAVLYPHGHHYIKRSLD----- 218
 DB 166 TEBAPEETEEDKATTEE-----APAAETSKAT-----EBAKAEETSAATEAP 212
 QY 219 -----ASELAAKAHLAGKMPOLSYSTASDNN 249
 DB 213 KAEETKATEAPKTEEDDKYTEAPRAEETSKATEAPRAEETNKVETEAPAAE 272
 QY 250 TQSVANGSTSRANKSENIQSILK-----ELYDSPAQRYSESDGLVDPANIIISR 300
 DB 273 TNKAAEET--PAVEDTNKNSNNAOPSETERQVYDVANKDYKKSEVTEAKAELEKY 330
 QY 301 TPNGVALIPROGDHYHFIPTSKLAL-EKIKAMPISGTGVTNKPKNVSVLSLSLS 359
 DB 331 LPKDI-----SMLSNEIKKIALSEVLKETAN--KENAQPRATFRSVSNAR 375
 QY 360 NPSSILTSKELSSASDGYIFNPDIYEETATAVIRHGDHFHPIKPSNQIGOPTL----- 414
 DB 376 TTVNNTSATLRAAA-----ODVYTKGTGFTLHAGDIIHTTYKEEPNEGTILAFNT 428
 QY 415 ---PNNSIATP-----SPSLPINPSTSEKHEEDGYGPDANRIIAD---ES 455
 DB 429 NPNPNTCKALRYNDKIDENKDFITVIVPANNNOGNTTADQMGFVFTGNGQDFLNG 488

QY 456 GFVMSGHGNNHYFFKKDLTEQIKAAQNHLEETKSHGCLSLSHEDDYPCGNATEKMDL 515
 DB 489 GILROKGMANSKFKIDPAAYNVNMKVDKLDADKT--NLSIQIGAAKGYGTGFVNGAD- 545
 QY 516 DKIKIEKAGIKQKGVKRESIVYKKEK--NAIITPHGHHADPIDKKPVGIGSHS- 572
 DB 546 -----GVNNOVG-----QNALNTRDKRPVNTIITADNTTNLD-----GQFHQ 583
 QY 573 -----NYELKPREGVAKKRGKRYTGEELTNVNNLKNSTFNNQNTLANQKRVSF 625
 DB 584 RINDVYLAND-----AATSIITATYAGTKATKATTDGIDSKQYNFLIISSHQNKY 636
 QY 636 S---PPELEKKGIMMLVKLTTPDQKYLEKVSAGVFGSGVNIANFELDQYLPQGTQK 682
 DB 637 SNGIRTNLEGS-----VTITTPQADLIDV--ETTKQPIPKHITREEDPILERS--- 684
 QY 683 YTIASNDPVEV---SYDS--FTYPTSLATYKASGTIYPPHAGDTYLRVNDQFVPGK 736
 DB 685 -----PDVIVQKGEDEKTTTPT---KVDPT-----GDVERGEPTTEVTN 725
 QY 737 -TDALVRY-----DEFGNAYLENNYKGEIKLPIPKLNOGTTTAGNKIIPVTEM 786
 DB 726 PVDEIVHFTPEEYPOGHKDEFPN-----LPI-----DQTEVPK----- 760
 QY 787 ANAYLDNSTYIVVPILEK-----ENQDKPSILPQKRNKAQENSKIDETVEEP 837
 DB 761 -KPGIKNPETGVVPPVDVYTKHGPKAGEPEVTKKEIPEFKKEEPPDLKGEERY--- 816
 QY 838 KTSVEYKEKLSGTGNSNSLSLEY---PVDPQOEVAFAEYQKLE-----N 886
 DB 817 -TOEGGTGKETTPTTPTINPLGEEKVGEPEPTTEVKEPEVDITQGGEEVPOGHKDEPD 875
 QY 887 VLENNQGTIELVLP-----SGEVIKNNADFTGEAPQGN----- 920
 DB 876 PVLPIDGTE--VPGKPGIKNPETGVVPPVDVYTKHGPKAGEPEVTKKEIPEFKKREF 933
 QY 921 -----GENKPSBNGRVSTGV-----ENQPT---ENKPADSLPEAPNPK 956
 DB 934 NDLKPGKEKTVQEGGTGKTTTPTTINPLGEEKVGEPEPTTEVKEPEVDITQGGEE 993
 QY 957 -----PAKPNSTON-----GNLPE--GNVSSDMLDPALEAPVNDVQOKLE 999
 DB 994 VPOGHDEPDPLPIDGTEVPKPGIKNPETGVVPPVDVYTKHGPKAGEPEVTKKE 1052

RESULT 7
 ID IG44_HAEIN STANDARD; PRT; 1849 AA.
 AC P45386;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE IMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IG41 PROTEASE).
 GN IGA.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NH1 HK61;
 RX MEDLINE-92234949; PubMed-1373717;
 RA Poulsen K., Reinholdt J., Killian M.:
 "A comparative genetic study of serologically distinct Haemophilus
 influenzae type 1 immunoglobulin A proteases."
 RL J. Bacteriol. 174:2913-2921(1992).
 CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
 CC -1- PRODUCING INTACT FC AND FAB FRAGMENTS.
 CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
 CC CERTAIN PRO-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
 CC SUBSTRATES ARE KNOWN.
 CC -1- SUBCELLULAR LOCATION: SECRETED.

Query Match	3.88;	Score 203.5;	DB 1;	Length 1849;
Best Local Similarity	19.38;	Pred. No. 0.069;		
Matches 239;	Conservative 153;	Mismatches 449;	Indels 397;	Gaps 59

```

0Y 28 HSHOEKONNRASYDGSOSKSELTPTDYOSAREIGD-----ABQIVIKITQ 77
Db 125 HDVYSEERK-----YTVVEKNPPEVWISFTTKEDDACKREDDYMPRLDFVEYEVADI 181
0Y 78 GYVSHGDHYHYNGKVPYDALF-----SEELIMDDPNOL-----KDAQ---YVNEVKEG 125
Db 182 EASTANNNGE--YNSDKPAPFVRLGSGGOFYFKKGSRRQLLTERDKGNLLRMNDVGG 240
0Y 126 YIIKDYGRYYVKKMAHADNRTDEIRKQOEHYKONEKYNNAVAARSGRRITTDNG 185
Db 241 DNUELYGNAYTY-----GIAGTYKYWHEN-----NGLIGFNSKEEHSDDPG 283
0Y 186 TY----FNPADIIEDTGNAYTY--PHGGHYHYTPKSDL-----SASELAAKAKHLAGR 232
Db 284 ILSODPLNYAVALGDSGSPFLFYDBREKGMFLGSYDFMAGYNNKSMQWOMNYIKHEFAEK 343
0Y 233 NNQPSQLSYSS--TADNNTQSY--AKGSTSKRANKSENISLKLLEYDSPAQR----- 284
Db 344 IYO-----OYSAGSLGSMYQTYMTQGSTITTTGGEPFLSYDLLTGKDKPBNHGKSTITLKG 399
0Y 285 -----SESDGLFEDPAKILSRPN-----GVAIPHQD-----HYFIFYSKLS 322
Db 400 SGLTLNNHIDQAGAGLEFEGDYEVKGRSDSTYTKGAGVSAVDGTYVMKYKNPNFYDRLA 459
0Y 323 -----ALEEK-----IARMVPISGTSVSTNKK-- 346
Db 460 KIGKGTIVVEEKGKNEGILLKYGDGVILLKORADANNKYVQASQVAYSGRSTLYVINDRQ 519
0Y 347 --PNEVYSSL--GSLSSPSSP-----TTSELSASQGYFENPKD 383
Db 520 VDPNSIYGFPGGRDLNGLNGBLTPHININIDGARVYVHNHNTNINITITGESLITNP-- 577
0Y 384 IVEEATATYIYRHGDFH-----YTPSNQIGOPTLPNNSIATPSPLPINPG 431
Db 578 ---MTTYSYNEADDDHPLRIRSIPIRYOLTFENDNR--SYTLLKKG--ASTRSELPQNSG 631
0Y 432 TSHHEHEDGYGFD-----ANRIIAEBSGFWVSHGDNHYEFPKDLTFOQIRAAQHL 485
Db 632 ESNNENMLYMGTSDEAKRNVNHNINNEKMGF-----KRYF-----GSEETKATQON-- 677
0Y 486 EYVYTSHGDLDSLSHEODY--PGNAEMKDLDKKIEEKIAGIMKQYGVKRESIYVNER 543

```

Db	678	GLKLTAVNG	-----KSDQNRFLITGGTINLNDL	-----	NYEK	705
Qy	544	NALITPEGDHHADPIDBEKPVGIG	-----HSHNSTELKPEEGVAKR	-----	EGN	589
Db	710	GLTLFISGRPTFARDI	-----AGISSTKPKDHFTEENNEVVEDWLNRFKATTAHVTCN	-----		764
Qy	590	KVYTGGEELTNVNLKSTNNQNFILANGQKRVSSFPPELEKLGIMLVK	-----			642
Db	765	ASLYSGR	-----NVANNTSITASN	-----NAQVHIGY	-----RTGDTVCVRSDYNGY	807
Qy	643	LITPDGKVLKSGKVFGEQY	-----GNI	-----ANFELDDPYLPQGTFRYTTIASKDYPVSDV		696
Db	808	VYCHNSNLSEKALNSFNNTNLKGVNVLLENASFLTKANLFG	-----	-----TLQISGTGSQVN	-----	860
Qy	697	GFTVPTSLAYKMASQTIFFPHAGDYTLRVNPQFAVPRKGTDLAVRPFDEFIGNA	-----			751
Db	861	-----LKENSWHMLIGNSVMNOLNLNGHILHLMNQANAKTYTNTLTVNLSIGNSFTYV				916
Qy	752	-----YLENNYKVGELKLP	-----PKLNQST	-----TRTGNKIPVTFMANVL		791
Db	917	VDETNNKSNKYVVKMSANGNFTLVQVADKTGEPHNNHLLTFDASMATRNMLEVT	-----LANGSV			975
Qy	792	DN	-----OSTYIYEVFTLEKENQ	-----DKRSILPQFRNKADQEN		827
Db	976	DGAKMYKILRVNNGRITDLYNPEVEKRNQIYDTNITTPNDIQADAPS	-----	-----AASN		1022
Qy	828	SLIDKVEEP	-----	-----KTSKRYEKELSETGNS		855
Db	1027	NEEIRAVTEPVPAPPAATESIASSEQETRPALFADPAVEENTVANSSTAPKSDTAQOT				1088
Qy	856	SNSTLEEPYTPDPOEKVAFABEYSGKLLNVLFNQDGIETLYLBSGEYIKKNMADF	-----			912
Db	1087	ENPNSSEVPS	-----ETTEVVAEHPPOENETAKNQEATEPTLPQNGEVAKEDQPTVEAN			1141
Qy	913	-----TGGAPOGNGENPSENSKATSTGYENQST	-----ENKPADSLPEAPNEKPY	-----	K	959
Db	1142	TQTNENQOSEKTEETO	-----TAEKSEPTESYTSVSENOPEKTVSOSTEDKVVYKEEK			1196
Qy	960	PENSTDNGMLNPEGNVGSDP	-----MLDPLALEAPVDPQVE			996
Db	1197	AKVETEETOKAPQVTSKEPKQALEPAPAEVPTDINAE				1234
RESULT 8						
MSPL_PLAF3		STANDARD;		PRT;	1682	AA.
AC	01-FEB-1991	(Rel. 17, Created)				
DT	01-NOV-1997	(Rel. 35, Last sequence update)				
DT	01-NOV-1997	(Rel. 35, Last annotation update)				
DE	MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)					
DE	(PUMSA) (P190)					
GN	MSF-1.					
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.					
OX	NCBI_TaxID=5834;					
RN	[1]					
RP	SEQUENCE OF 1-1061 FROM N.A.					
RX	MEDLINE=88166657; PubMed=3373688;					
RA	Cetta U., Rotmann D., Matile H., Reber-Lisse R.,					
RT	"A naturally occurring gene encoding the major surface antigen					
RL	precursor p190 of Plasmodium falciparum lacks tripeptide repeats."					
RL	EMBO J. 6:4137-4142(1987).					
RN	[2]					
RP	SEQUENCE OF 1032-1682 FROM N.A.					
RX	MEDLINE=95354793; PubMed=7628566;					
RA	Tolle R., Bujard H., Cooper J.A.;					
RT	"Plasmodium falciparum: variations within the C-terminal region of					
RL	merozoite surface antigen-1."					
RL	Exp. Parasitol. 81:47-54(1995).					
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR					
CC	(PORENTIAL)					
CC	-1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42					

KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MERIZOTES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.slb.ch/announce/>
CC or send an email to license@isb.slb.ch).

DR EMBL: M35727; AAA29715.1;
DR EMBL: Y00087; CA868280.1;
DR EMBL: 235326; CA84555.1;
PIR: S06286; S06286.
InterPro: IPR000561;
DR Pfam: PF00008; EGF, 1;
DR Malaria: Merizote; Polyprotein; Repeat; signal; glycoprotein;
KW Transmembrane; GPI-anchor.

FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1682 MEMBRANE ANCHOR (POTENTIAL).
FT TRANSMEM 1666 1682
FT CARBOHYD 233 233 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82A1E15948CAD6 CRC64;

Query Match 3.5%; Score 189; DB 1; Length 1682;

Best Local Similarity 20.0%; Pred. No. 0.28; Mismatches 428; Indels 362; Gaps 64;
Matches 242; Conservative 177;

QY 1 MKRSKRYIAAGSAIVISLCAIYALNOHRSQEN---KDNRRVSYDQSGOSKSENTP 56
492 MKRNNF---NKDYVDKIPSAKYTYVEKORNNKSSSSSYVNV---QKLKALSYLE 545
QY 57 DOVSOREGIGAEQIVKIDQGYVTSBGHYHYNGKVPYDA---LFSEEL----- 104
546 D-YSLRKG-----SEKDNHYTLKTGLEADIKKTEIKSEKILE 588
QY 105 ---LAKDPYOLKADQYN-EVKGQYI-KYDG---KYVYTKDAHADNVT----- 149
589 KNEKGLHSANASILEVSDIYKLOVRYLLKIKIEDLKIELEFKMQLKDSIHVPNIKP 648
QY 150 ---KDEINRQK---EHYKDNKRYNSVA---VASQGRYTTNDGYVF 188
649 QNRPETLYLVKKEVDKLEPIPKYDMKQDAVLSITQPLVASE---TTEDS--- 702
QY 189 NPADIIDETGNAYIVPHGHHYIPKSDLSASELAAKAHLAGKNMPSQLSTSSASDN 248
703 ---GSHTHLSQSGELEVETEVEYGHVTTITLPPREYVENSIEHKNSND- 753
QY 249 NTQSVANGSKSPANK-----SENLSLKEIYD-SFSARYSQDGLVF 292
754 NSQALFTYVLYLKLDELFRSYCHKYILVNSNMDKLEVNULPEENELSCDRL-- 811
QY 293 DPAKISRTNGVAIPHG-----DHYHFIPYK---LSALEEK-IAMV----- 332
812 ---DLNINONNIPAMSLDSNANDQHLFEIYOKEMIIYLLKLEENHIKILLPEK 868
QY 333 PISGIGSTVST-NAKPREVYVSLIGSLSSPSLITSELSASDGYTFNPDYIEETATA 391
869 QITSTSTSPGNTVTWASATHSNOSNOOSNASS---TWTONG-----VAVSSGPA 918
QY 392 YIVHGHFHTIPKSNQI-GQPLPN--NSLATSPSLPILNPGSHKHEHDGIGFDANR 448

DB 919 VEEESHDPLTVLISNDLGIIVSLNKGRTVPPN-LTIST-TEKENYEYLKI-MIP 975
QY 449 IIAEDSGFVMSGHNNHFFKDLTEBOIKAAQKLEEVKTS-----HN-----GLDS 497
DB 976 IINDDIKQVKNNSK-----VTGLETOKNALDEIKKLTQLSFDLYNNKYLKLD 1030
QY 498 LSHEDDYGNKAKEMD---LDKKIEKTAGIKOYGVARE-SIVNKKEMALI----- 547
DB 1031 LNKKELEGDDQKIKLLKLEQESKINSLNPNPNVONFVFNKKKEAIAETENT 1090
QY 548 ---YPGDHHADPIDBKPVGI--GHSNNTLFRPEBVAKKEG 589
DB 1091 LENTKILLNKGKLVYNGE---SSPLKTSFVSJOTEDNVAANKLFRSLKIDKLAND 1147
QY 590 KYTGTGE-----LTNVNLIKSTFNNOFTLNGOKRVSPPEPELEKIGIMLV 641
DB 1148 NMLGKRKLSPLSSGHHLLTELEK-EVINKNT-----GNSPSENNK--VNEAL 1195
QY 642 KL---ITPDGKVLKESGKVFGEVGNIANFELDQYLPQGTFRYTI---ASKDYPEV 693
DB 1196 KSEYENLPKATVTVTP-----POPDVPSPLSVRVSGSGSKTEETQI 1240
QY 694 SYDGTFTVPSLAVKNAQOTIYFPAAGTYLVNPOFAVPGTD-----ALRV 743
DB 1241 PISG-----SLTELOOVOLONDEEDSLVPLIFGSESDNDEYLDQVVTGEALST 1294
QY 744 PDEPHNAVLENRYKGEIKLPKLNQGTTPRAGNIP---VTFMANA----- 789
DB 1295 MONILSG--PENEDVIYIK-PL---AGYRSLKQIENITTFNLINDILNSRLKR 1347
QY 790 ---YLD-----NOSTYIYEPILKEKNOTKPSILPOFKRNK-----AOE 826
DB 1348 KYLDVLDESIDMOKFIHSSHEIIE-DSFKLNSQKNTLLKSYKIKESVENDIDFAOE 1406
QY 827 NSKIDKRV---EERTSEKVEKELSETGNSSTSLTEVTPVPOVQAKAFKA----- 877
DB 1407 GISYERVLAKYKDDLESIKVIKE--EKEFPSSPTTPSPAKDEOKSEKFPFLT 1463
QY 878 ---ESYKMKLENTLFNN-----DGTIELYPSGEV 904
DB 1464 NIETLYNNLVKIDYILINKAKINDONVEKDEAHYKTYLSDKAIDKIDLF----- 1517
QY 905 IKKNADFTGEAPQGNENKPSSENGK-VSTGYVENOPT-----ENKPADSPEAPNE 955
DB 1518 ---KNPIDEFAIKLINDTKDKMLGKILSTGLVGNPNFTIISLBEKFDMDNISQHC 1575
QY 956 -KPEKPENS 963
DB 1576 VKKQCPQNS 1584

RESULT 9
P3K1.DICDI STANDARD; PRT; 1570 AA.
ID P3K1.DICDI
AC P54673;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PHOSPHATIDYLINOSITOL 3-KINASE 1 (EC 2.7.1.137) (PI3-KINASE)
DE (PTDINS-3-KINASE) (PI3K).
GN PIKA OR PIK1.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-AK3;
RX MEDLINE=96009592; Pubmed=7565716;
RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT A phosphatidylinositol (PI) kinase gene family in Dictyostellium
RT discoideum: biological roles of putative mammalian p110 and yeast
Vps34p P1 3-kinase homologs during growth and development.*;

DR	EMBL: U23476; AAA85721.1; -
DR	DictyDb; DD01099; p1Ka.
DR	InterPro; IPR000341; -
DR	InterPro; IPR000403; -
DR	InterPro; IPR001263; -
DR	InterPro; IPR002420; -
DR	p1fam; PF00792; PI3K_C2.1.
DR	p1fam; PF00794; PI3K_Rbd.1.
DR	p1fam; PF00613; p13Ka; 1.
DR	p1fam; PF00454; p13, p14_kinase; 1.
DR	PROSITE; PS00915; P13_4_KINASE.1; 1.
DR	PROSITE; PS00916; P13_4_KINASE.2; 1.
DR	PROSITE; PS50290; p13_4_KINASE.3; 1.
-	Transferase; Kinase; Multigene family.
-	DOMAIN 41 48
FT	DOMAIN 59 73
FT	DOMAIN 161 168
FT	DOMAIN 294 304
FT	DOMAIN 308 324
FT	DOMAIN 413 424
FT	DOMAIN 503 510
FT	DOMAIN 570 579
FT	DOMAIN 821 828
FT	DOMAIN 831 836
FT	DOMAIN 1309 1570
SEQUENCE	1570 AA; 178374 MW; 55B678B7B34D783 CRC64;

[illegible]

QY	469	-----	FKKDLTEBQI	-----	KAQK	483
Db	459	KNTNPNOLSTSSQ	FFKTLTSLN	ETPIDCKINDINDPTDAPSDLSASASSSSFTTSSOS		518
QY	484	-----	HLIEKTSHNGCLDSLSHEDQYPGNAKEM	-----	KDLCKTEEKIAGIKM	528
Db	519	LNVGSLRKAKTSP	FNLIPLFM	-----	PQSKRIIQVYGSDPIENLAKERRITSDLPFN	571
QY	529	-----	QYGV	-----	KRESIVVN	548
Db	572	NNNNNNNNCKY	GADSYLLIDPNDPMERSL	VLNKSDYILDRAGGLPKLVIEKSTIL		630
QY	549	PHGDHHDAPDIDENK	PVGIGSHSNVLEFK	-----	PEEGVAKK	587
Db	631	-----	DDSPDSELS	-----	SEYELIRRLIGDTMGEVEYFRVTSRLREALP	677
QY	588	-----	GKRVYTG	-----	EELTNVNLKNTS	617
Db	678	LKSGIOSTLIVRLSP	PIPIVGNKILISLIFLP	ITVQTKLIDLEMETADQFNRLFT	---	735
QY	618	NGQKRVSFSPPELEK	KGIMDLVKLITPDQVLEKRVSGFVPEGVGNIANEE	LDQPL		676
Db	736	-----	KMYSHLPN	-----	VNSHFLIKVYGSDDFHGHDRFTFESINHHI	778
QY	677	PGQTKYTIASKDYPE	SYDGTFFVPTSLAVKASQITFFPHAGDTYLRVNPQAVPRG			736
Db	779	QGTQDQLILQRPPELD	PO	-----	PF	809
QY	737	TDALVRVDEFGNAYLE	NN	-----	YVGSIRKLPIKLNQCTR	775
Db	810	L	-----	ILDHCSNAINCNNNTNNTNNNNINFDNDQITHSIRIKKPRFVYWGSTR		864
QY	776	TAGNKIPPTFANNAYL	DNOSTYIYEVPILEK	-----	ENQDCKPSILPOFR	820
Db	865	-----	IPLSCKID	-----	DSSSVIATISLVHGICBFSKAFQPIIPPPFALETISVDMC	916
QY	821	-----	RNKAQNSKLDERVEEPT	SEKVEKEKISLSTGENSTNSTLEEPTVDPVOEKVAK		875
Db	917	EWLFTNTIDYSNLE	VDARL	-----	SISVYSANETVDVDE	951
QY	876	FAESYGMLENVLE	PNMDGTIELYLPSEGVTKKMMADFTGAPQNGENK	-----	PSENGKYST	933
Db	952	-----	INLDEATRKLLP	LG-WINWITTFKYQLRGVYELSLMPSDSN	PL	997
QY	934	GIVENOPTENKPAD	SLPE	-----	APNEKPYKPNSTDNGMLNPREGNVSGDMLPALTEE	987
Db	998	GTCGNSPSSQSVGL	TEFEFENPLPVLPFRKTRFTFSVSVIQTNPPTNINSNKR	-----	FFQ	1056
QY	988	APAVDPVOE	KLEKFTASVGLGDSVTF			1014
Db	1057	ITALDPLSLDKOEKYNOLMTLRHSILF				1084
RESULT 10						
SC16_YEAST						
AC	SC16_YEAST	STANDARD		PRT	2194	AA
AC	P48415					
DT	01-FEB-1996	(Rel. 33, Created)				
DT	01-FEB-1996	(Rel. 33, Last sequence update)				
DT	15-DEC-1998	(Rel. 37, Last annotation update)				
DE	MULTIDOMAIN VESICLE COAT PROTEIN.					
GN	SEC16 OR YPL085W.					
OS	Saccharomyces cerevisiae (Baker's yeast).					
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotini; Saccharomycetes;					
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.					
OX	NCBI_TaxID=4932;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	MEDLINE=96011704; PubMed=7593161;					
RA	Ebenshade P., Glimeo R.E., Holzmancher E., Teng P., Kaiser C.A.;					
RT	"Seest SEC16 gene encodes a multidomain vesicle coat protein that					
RT	interacts with Sec23p."					
RL	J. Cell Biol. 131:311-324(1995).					

CC -1- FUNCTION: INVOLVED IN THE BUDDING OF TRANSPORT VESICLE FROM THE
 CC ENOPLASMIC RETICULUM. THE C-TERMINAL INTERACTS WITH SEC23 AND
 CC WITH THE CYTOSOLIC DOMAIN OF SED4. COULD THEREFORE BE A
 CC CONSTITUENT OF COP1 VESICLE COAT. N-TERMINAL OVEREXPRESSION
 CC CAUSES A LETHAL SECRETION DEFECT.
 CC -1- SUBCELLULAR LOCATION: ON THE ENOPLASMIC RETICULUM AND ON VESICLES
 CC WHICH BUD FROM IT.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).

DR EMBL: U23819; AAC49088.1;
 DR SGD: S0006006; SPC16.
 KM Transport; Protein transport; Golgi stack; Endoplasmic reticulum.
 FT MUTAGEN 1058 L->S; IN SEC16-4; TS ACCUMULATION OF ER
 FT MUTAGEN 1083 1083 MEMBRANES.
 FT L->P; IN SEC16-3; TS ACCUMULATION OF ER
 FT MEMBRANES.
 FT MUTAGEN 1088 1088 L->P; IN SEC16-2; TS ACCUMULATION OF ER
 FT MEMBRANES.
 FT MUTAGEN 1230 1230 M->R; IN SEC16-1; TS ACCUMULATION OF ER
 FT MEMBRANES.
 SO SEQUENCE 2194 AA; 241613 MW; BB1E02D2AD4683E3 CRC64;

Query Match 3.4%; Score 185; DB 1; Length 2194;
 Best Local Similarity 21.6%; Pred. No. 0.61;
 Matches 223; Conservative 124; Mismatches 418; Indels 266; Gaps 54;

QY 92 GKVYDALESEELMMDPYOLKADIVNEVKG-----YIKVDGK--- 133
 DB 1282 GNP-----LSNEP--VADSDVIFESTGNMTFESILMDEIYEIYFSDPRFKGF 1330
 QY 134 -----YYVYLKDAHAADNVTK--DEINQ-KOEHYNDKENV-----SNVAVARSO 177
 DB 1331 SSILPKIYHASILQEGNLISLTQTKYTLSSVYRKLPKRDILTINLTRELSVVASRLSE 1390
 DB 178 GRITTDGVVNP-----DIEDGNATYVPHG-----HYHYIKPSDL 217
 DB 1391 ----SWTGMIAKRLSSVWGQIDKRSFKYI--GGDDIDALNKNDKKVYDFGFGSSA 1443
 QY 218 SASELAAAHLAGKNMPSOLSYSTASDNMTQSAKST--SKPANKSENLQSLKEL 275
 DB 1444 NSSTVDLTQTFPFOAQVTSQSYVDTALLNNAHNPVSHVLSKTSNVS-----KGL 1496
 QY 276 YDPSAORTSESDGLVFDPAKIIKTPNGVAIPH-----GDHYHFIYPSKLSALEE 326
 DB 1497 VEANLPYTHRIGSLQSPQR-IHNTQFAALBPOMASLRKRYRDOHTNEKALKSOQILKE 1555
 QY 327 KIARMPISGSGVSTNAKPNVSVSSLSNPSLITSKLSASASOYIINPDIYE 386
 DB 1556 KSTAYTPQFQONHVPKSNVSNVPSLADFAPPP-----KLGVTYSNTVSSP-DLYR 1607
 QY 387 ETATAYVVRGHDHFIYPKNSNOIGOPTLPNNS-----LATPS-PSLIPNPTS--HEKHEE 439
 DB 1608 RES---TISGSEFLPPRK--IGVTPKANSOGSLMTSSVYKLPIDPVVQVHERGYN 1661
 QY 440 DGYGFDAARIIADES--GFVMSHGHNHYFFKKDLTEQIKAAQKHLVEYKTSNGLDS 497
 DB 1662 DFGKHSQKAMPDESHSHDNSNADONTLKDSADVDYDEMTDIEGPGFNDVY--NLLPM 1718
 QY 498 LSSHEDYPRNAEKMDLKIETKIAIGIKQGVKRESIVYKREKAAIITYPHGDHHD 557
 DB 1719 EPNHQPSTVNPLO-----TISDDIOPIL-QTNVEVRGTDASQWENSL--PSIENERS- 1768
 QY 558 PIDHKKVVGIGHSNHYELKPEEGVAKKREKNKYTGEEELTNVNLKNSTFNNQNTLA 617
 DB 1769 --SEEPENISKASNS--ATLPSTGGISLE-NRPITODE--NSISEVQOSTY-----LP 1815

QY 618 NGOKRVSEFPELEKLGINKLVLTIPDGNVLEKVGATPGEVGNIA----- 667
 DB 1816 AG-----SISMAKPIQOVDPVPRNVNKKASKLVEQNHAPPKRSTDA 1859
 QY 668 --NEFLDPIYLGQFKTKTASKDVPEVSYDTFFVPSIAYKMASO-----TIFYPP- 718
 DB 1860 KKNY---SPYVP-----QSTVMSADDE--STIIKTPAIARPHQAHASPSQTFPLV 1908
 QY 719 -HAGDTYLRVNOFAVPGTDALVAVDFEFGNATLNNYKGEIKLPKLN----- 770
 DB 1909 NQANET-----ASFELSESTSOA-----QSNQNVASENFS-----PIKALVEYKDTF 1952
 QY 771 QGTTTAGNR-----IPYTFMANAYL-----DNOSTYIYEVPILEKENQDTPSL 816
 DB 1953 OPTIKASINOVRAKPLESDAKTDNDYEDSDDDNNTDPAKRNKEKKNVNMKEK 2012
 QY 817 PQFRNRKAQENSKLDEK-----VEEPTSEKVEKEKISEFGNSTSNTELEVTPDY 869
 DB 2013 P-----SNKDIDKSNQWFGWLKIDGDKVYKAKIGHNNTLYYDEKLRWVRKAT 2064
 QY 870 QEKVAKFASVGMKLENTLVNMD-----GTIELYLP--SGEVIRKNNADFTGE-- 915
 DB 2065 EEKOKIIESSAPPPPIYKRRDGGPKTKRSGPPIINSLPVAHAYVLPNN--PTGEPD 2122
 QY 916 -----APQG-NGENKPSNGKVS--TGTVENQPEENKPADSLPEAPNEKP--VKPENST 964
 DB 2123 PKTSPSPGPPPNNSPSPSPISRLSGVNLASKANGDLDLSLAGGPRPASTRAKKT 2182
 QY 965 DNGMLNPEGNV 975
 DB 2183 ARGVAVMDNI 2193

RESULT 11
 SP2 YEAST
 ID SP2 YEAST STANDARD; PRT; 1466 AA.
 AC P23201;
 DT 01-NOV-1991 (Rel. 20; Created)
 DT 01-NOV-1991 (Rel. 20; Last sequence update)
 DT 30-MAY-2000 (Rel. 39; Last annotation update)
 DE SP2 PROTEIN
 GN SP2 OR PEA1 OR YIL021W OR YIL209.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8288C;
 RX MEDLINE=91009481; PubMed=2211820;
 RA Gehring S., Snyder M.,
 RT The SP2 gene of Saccharomyces cerevisiae is important for
 RT pheromone-induced morphogenesis and efficient mating.
 RT J. Cell Biol. 111:1451-1464(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8288C;
 RA Fumelle B., Goffeau A.,
 RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN PHEROMONE-INDUCED MORPHOGENESIS AND
 CC EFFICIENT MATING. PERHAPS AS A CYTOSKELETAL PROTEIN.
 CC -1- SUBCELLULAR LOCATION: LOCALIZES A SHARP PATCH AT THE SHMO TIP
 CC (MATING PROJECTION) WHICH IS THE SITE OF POLARIZED CELL GROWTH.
 CC -1- SIMILARITY: TO PROTEINS THAT FORM COILED-COIL STRUCTURES.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).

Query Match	3.48;	Score 184.5;	DB 1;	Length 1466;
Best Local Similarity	18.28;	Pred. NO. 0.37;		
Matches 204; Conservative	163;	Mismatches 386;	Indels 369;	Gaps 49

```

Db      660 SNLSLPRKLT-----LPQAKIGVYIPSENV-----PIKILMTE 696
Qy      439 EDGTGDANRIIAEDSGTASHSGDHNYFFPKDLTEE-QIKAKOLEEYKTSNHLDS 497
Db      697 ED-----NK-----RSDITNEISVYKPTSSIAADLKQFEDSSSK 729
Qy      498 LSHBEDDYGNKAKENKDDJDKIIEEKIAGIMQYGVKKESTIYVNRKNAIITPPGCHHAD 557
Db      730 KSSPKKN--PIAKEEDRKPKLSNFKITSMNDVSTDDSSSDGN-----ENDDAD 776
Qy      538 PIDEHPVIGISHSHSNTELFKPEEGVAKKEGNKYTGEEL-----TNNVLNKTFFNONF 614
Db      717 DDDDFYAL-----KQMKRESKIEKRNNDKPLANTYELDLSPSPSVYI 823
Qy      615 TLANGOKRVYSFEPPELEKKLGIMLVKLTTPD-----GRVLKYSKAVFGEGVN 665
Db      824 ESEPSIKETTS--EKSSEMPSSSLPRLYEDVEPESEMPKASVSYAKKNFOEPLGN 880
Qy      666 IANFELDQ-----PYLPGQFFKTIASK-----688
Db      881 VSPDPTQIVNSLGTGTGAAGPESDSRYESPGMTGQIKSLIMAGKVVPEADSRVSPGM 940
Qy      689 --DYEVSYDGTFFVPSL-----AYKASOTIYPPHAGDTFLRWQFAVPKGTDALY 741
Db      941 KQKISLSLGTGTGTTAOSIKSPEARLTAS-----SEGVKISPPMV--RESESL 990
Qy      742 RVFDFHGNAYLENNKVGELKILPIPLNOGTTTACNKPITYFMNAVILNOSTYIYEV 801
Db      991 AV-----CNTPSPNV--TVKMSPTL-KGNVYSEQIEIRDDLSSEPIEN-----1032
Qy      802 PLEENQTDPSLLPQFKRKAQENKLDENYVEPRTSKVLEKLSGTGS-----T 855
Db      1033 -----VDPKVLKTIYVPEKAVNRRGSPKSEVETKPSATLKSLPPEPNQIYSPELA 1084
Qy      856 SNSTL-----EEVPIYVDVVOE-----RVAKFAESYGAK 883
Db      1085 KNSPLAPIKKAVELRENNKPTETITTSVEPTNKDANTSMFDLNRITKREDEDFDVR 1144
Qy      884 LEVV---LFNMDSITELY-LPSGEVYIKKMAFTGEAPQNGENKPSNGFVSTGYE- 937
Db      1145 NNNIGITGAYVTKTKIDYHKIPYDRKAKSAEAVHTSEEDIDSNN---VNGKRDAQCHI 1201
Qy      938 -----NQPENKRPADSLDEAP--NEKPYKPENSIDNGLN 970
Db      1202 TERKHAIVNPTENSQVAKTSHSPFLNSKPYOVENSSNGIN 1243

RESULT 12
BDD3_YEAST STANDARD: PRT, 1636 AA.
AC BDD3_YEAST P25557, P87007;
AC P25558, P25557, P87007;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE BDD SITE SELECTION PROTEIN BDD3.
DE BDD3 OR YCLO14W OR YC14W/YC11W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RN RA Chant J.S.;
RN RL Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RP RA Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RP RL Submitted (MAR-1992) to the EMBL/Genbank/DBJ databases.
RN [3]
RN REVISIONS.
RN RA Gromadka R.;
RN RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.

```

CC - FUNCTION: CO-ASSEMBLES WITH BUD4 AT BUD SITES. BUD4 AND BUD3 MAY
 CC COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
 CC DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
 CC ESTABLISHING THE AXIAL BUDDING PATTERN IN G1
 CC - CAUTION: REF. 3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1338
 CC ONWARD AND IS SHORTER (1367 AA) DUE TO A FRAMESHIFT
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).

EMBL: 017580; AAA86315.1; ALT_FRAME.
 DR PIR: X59720; CAA42346.1; ALT_FRAME.
 DR PIR: S19341; S19341.
 DR PIR: S19340; S19340.
 DR SGD: S0000520; BUD3.
 FT CONFLICT 285 Q -> P (IN REF. 3).
 FT CONFLICT 407 K -> N (IN REF. 3).
 FT SEQUENCE 1636 AA; 184717 MW; 9E4E6BA5C3A3F69 CRC64;

Query Match 3.3%; Score 179.5; DB 1; Length 1636;
 Best Local Similarity 18.7%; Pred. No. 0.72;
 Matches 186; Conservative 135; Mismatches 327; Indels 347; Gaps 47;

91 NGVPPDALFSEELMKDPYOLKADIVNEVGGYIIKVDGKYYVYLKDAHADVNRK 150
 722 NSSMADLLCVNENLKNHEHOLEV----- 747
 151 DEINROKOEHVKNENKVNVAVA-----RSOGRTYNDGYVNPADIIETGNA 200
 748 -----KHPSTDEHRAVNSKLSGASDPDAHEKKRSYGITTRSTSLDK----- 733
 201 YIVPHGGHYHYPKSDLSASELAARAHLAGKMOPOSLSYSTASDNTQSAKSTS- 259
 794 -SPSGDMSVTKET-----KELLPVKPKSKSKPKREIOKTKTKMAK 835
 260 -----KPAKSENLSLKELDSPSAORVSSDGLVPAPATISPTPGVAIPHGDH 313
 836 AEHIEKKPKPKGFGVGLNVGSKSKSPSVQVQV-----PKTISOHPPS----- 883
 314 HPIPKSLALEKIAWPIISGTVSTNAKPNVYSSLSISNPSILTSKELSA 373
 884 ---PVKPKPTSEK-----SSPKRAVSSPKTKKSTSEST-KESQTA 922
 374 -----SDGYIFNPKDIVERATAYIVRHGDHFHYIPKSNQIGQPTLPNSLATPS 423
 923 KSSIRAVEFKSDDLIGRPD-----VGNCAH-----POEWTRISVYRDKRYSYN 968
 424 PSLDIPNGTSHKEHEEDGYGFANDKRIIAEAE-SGFVASHGDHNYFFKKDL-TEEOI-K 479
 969 PSQVTEVTSNEKNVEP-----KADQSTKODNISNFADEVASASSYPRKILAEVDQITIGK 1024
 480 AAKR-----HLEEVTSHGSLDLSISHEODYPGNAKEMDLD-----KRI 519
 1025 ATNSSVHGNKELPDLEAVTTA-NKVTTSAGDDRIQOSFLLAAADVENISDDDEHON 1083
 520 EELK-----IMKQYGVRESIYVNNKKNATIIYPHGDHNDPIDEKRPVIGSHS 572
 1084 ESRVNDLFDGDFIPKHNKNQENI-NSSSN-LFPEG-----KVQOEK-CVSNMENT 1131
 573 NYELKPEEGVAKKEGKNTVYGEELVNVNLLKSTFNO-----NFLANOK 621
 1132 NISLKTNDASTLTKLSPQASKVLTENSNEIKDT--NNEKDAKDKLGLDGYDKETAK 1189
 622 RVSEFPPE-----LEKKIGIMLVKLTIPDGKYLEVSGKVGEGVGNANFELDPOY 675
 1190 EIT---KKNFVEGITEKKEIFPTIPRLAPASKI-----NQRSPSY 1229

676 LPQGFKTTIAKDYPERVSYDQTFVPSLAKMASQITFFHAGDYLAVNPOFAVRK 735
 1230 I---ELFOGKRVVLDKHDHAYNKK-----RLASQSL-----SEGLKN----- 1265
 736 GIDALVAVDEHNGNAYLENNKVGEL-----KLPPKLNQSTTRAGNKIPVTFANAY 790
 1266 -TEEDAAITKSDODKAKERMTOISEVYTEMQPIP-----TILPKAH 1308
 791 LNOGTYIYVEPILEKENDTPSILPQPKRNKADENSLDKERPEPKSERVENKLS 850
 1309 LDDSG-----IEKSD-----DKFEEIEELKEKLSKGTGNDVON 1344
 851 TGNSTNSSTLEVPY-----VDPQEKVAKPAESYKMLEN-----VLNMD----- 892
 1345 NNPNSIPRIEKPAPKAVTQRTSPVRIIGTFEDT--RKTENGSPSDISFTYDHNDEPD 1402
 893 -GTIELTLPSEGVYKKNMADFGAEVQNGENKSPNGKSVTGTVENQTPENPADSLPE 951
 1403 KRLMELKFPSSODEIPDDR-YT-----PAE-----EPTAEFPVELRN 1439
 952 APNEKPVKPEK-SDNGMILNPEGVNSDP--MLD 982
 1440 TRSIVTTSNNKSTDDKL--SSGNIDQKPTELD 1472

RESULT 13
 IGK2_HAEIN STANDARD; PRT; 1702 AA.
 AC P4384;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DT 01-NOV-1995 (rel. 32, Last annotation update)
 DE IMMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGK1 PROTEASE).
 GN IGA.
 OS Haemophilus influenzae.
 OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_Taxid=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HK715 / SEROTYPE B;
 RX MEDLINE=92234949; PubMed=1373717;
 RA Poulsen K., Reinholdt J., Killian M.;
 RT "A comparative genetic study of serologically distinct Haemophilus
 influenzae type 1 immunoglobulin A1 proteases."
 RL J. Bacteriol. 174:2913-2921(1992).
 CC - FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
 CC - PRODUCING INTACT FC AND FAB FRAGMENTS.
 CC - CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
 CC CERTAIN PRO-1-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
 CC SUBSTRATES ARE KNOWN.
 CC - SUBCELLULAR LOCATION: SECRETED.
 CC - DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
 CC SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE)).
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).

EMBL: M87489; AAA24966.1; -
 DR MEROPS; S06.001;
 DR InterPro; IPR000710;
 DR PRINTS; PR00921; IASERPTASE.
 KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal.
 FT SIGNAL 1 25

FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
 FT PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
 FT ACT_SITE 288 1124 PROBABLE.
 FT DOMAIN 1109 1124 2 x 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
 FT REPEAT 1109 1116 K.
 FT REPEAT 1117 1124 2.
 SQ SEQUENCE 1702 AA: 186539 MW: 860F70D2667807A6 CRC64;

Query Match 33; Score 177; DB 1; Length 1702;
 Best Local Similarity 19.68; Pred. NO. 1;
 Matches 223; Conservative 144; Mismatches 430; Indels 338; Gaps 53;

QY 84 GDHYHYNGRPV-----DALP-----SEE-----LAKDP--NQ-LKAD-----IV 129
 DB 236 GDAYTYGAGTGYKVNHEHNGLLGFQNSKEHSDPKGILLSQDLTYNAYVAGDSGSLFVY 235
 QY 120 NEVKGCIIVQCKRYVYLKDAHADNVTKEIINROKQEHVQNEKVNVAVASOGR 179
 DB 296 DREKGMFL--GSL-----DFWAGYKKSQWEMNITKEPAK-----TVLDKDTAGS 341
 QY 180 YL-TNDGYVNPADIEDTGNAYIVPHGHYIIPKSDLSASFLAAKHLAKNNQPSQ 238
 DB 342 LIGSNQYKMN-----IGKTSVINGSESLNVDLP--SSQDTSKNNHRSV---T 390
 QY 239 LSTSTASDNN-----TQSAKSTKPAKSENLSLITKELTSPSAQ 282
 DB 391 LRSGSTLTNNNIDQAGGLFEFGDEYVGTSDTWTWAGVAVAGKTYVTVHPKSD 450
 QY 283 RYSE-SDGLVFPDAKISTKPNVAILPHDDHFIPIYSLSALEKRIARVPISGSTV 341
 DB 451 RLAKIGKGLI-----VEGGENKSLKVDGTVILQADANNKVAFAQGVIGRSTV 506
 QY 342 STN-----AKPNEVSSL--GSLSSNPSSLTSEKLSASADGYIFNKKDYEETATYTR 395
 DB 507 VANDKQVNPNSIYFEGFRGRDANGNNL--TEHINIDG-----ARLVN 551
 QY 396 HGDHFIYKPSNOIGQPLPNSLAT--PSPSLPIPGTSHE-----KHEEDG----- 441
 DB 552 H-----NTSKTSVITIGESLITDPTIIPYNDADDEENPYAFRIKGGOLYLWL 603
 QY 442 --YGPANAIIDESGFVMSGDHN--HFFPKDLTEQIKAAQKHLVEKTSHGGLSL 498
 DB 604 ENYTYALAKGASTSELPRKNSGSENNMLYMKTSDEAKRMVHMINNEM--NGFNXY 661
 QY 499 SREO-----DYPGNAKEMKDL-----DKTEE-----KIKGI 526
 DB 669 PEEEGKNNGNLNTFKKSEGNRLLGCTNLNGDLVKEGTLFLSGRPFHARDIAGI 721
 QY 527 MK-----QGVKRESIV-----NKEKNAIYPRGHHAD--PIDHKPV 565
 DB 722 SSTKDKQFAENNEVVEDWVNRFKATNLVNTNATLISGRVANITSNTATADAKV 781
 QY 566 GIGHSNVELEKPEBGVAKKGNV-----YTGEELNVNVLKNSFPNNQFTLANGO 620
 DB 782 HIGY-----KAGDVVCRSDYTG--YVCTDKLSDKLSNFSNATNATVGN 824
 QY 621 KRVSEFPPELEKRLGINMLVLLTPDG-----KYLEKVSQVFGVGNINAFELDQYL 676
 DB 825 VMLS-----GNANFVLGKANLFGTISGTSNOVRLTNSHMHLLTGSNNQNLNDGHIHL 880
 QY 677 PEGQFKYTIASKD---YEVSYDTGFTVPTSLAYKMASQITFFPHAGDTYLKVPQPAV 733
 DB 881 NAQNDANVTYNTLVNLSGNSFYLLTDSNKGQKVVVTSANGNFLQVADKIGE 940
 QY 734 PKGDALVRVDEPHGNAVLENNKTVGEIKLPRLNQGTRTIAQNKIPVTFMNAVLYDN 793
 DB 941 P--TKNETLTLD--SNA--TRNNNLVSLV-----GNYVLDGAMKYLKLVN 980
 QY 794 OSTIYVEPILEKENC-----DKPSI-----LPQF 819
 DB 981 NGRDVLNFEVKNQVDTNTITTPNNIADQVSPVSNNEELARVETPPVPPAPAPSE 1040

QY 820 KRNAQNSKLDER-----VEEPKTEK--VEKEKLSGNSGNSNS 858
 DB 1041 TTEVAENSOESTVKNEDODATETTAONGEVAEKPSPVKANTQNEVAQSSSETEET 1100
 QY 859 TLEEVPTVDVQ--EKVAKFALSTGKLENTLFFNNDGTEILYLPBGVYKKNMADFGEAP 917
 DB 1101 QTTEIKETAVKEKAKVKEKAKYK-----DEIQEAP 1136
 QY 918 QGGENKP-----SENGVSTGVENOP-----TENKPADSL--PE 951
 DB 1137 QMASSESPKQAKPAKPEVSTDTKVEETQVOAQOTOSTVAAAEATSPNSKPALETPSE 1196
 QY 952 APKEKPKP-----ENSDNGNLNPGANGSPMDPALEAPAV-----DPOKLE 999
 DB 1197 KINAEPVTPVSKNQTENTDQPTEREKTAKEVTEKTOEPPQVQASQAPQOSE 1251

RESULT 14
 HMW1_MYGE
 ID HMW1_MYGE STANDARD; PRT; 1139 AA.
 AC 049413; 049365;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CYTADHERENCE HIGH MOLECULAR WEIGHT PROTEIN 1 (CYTADHERENCE ACCESSORY
 GN HMW1 OR MG312.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1].
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RC MEDLINE-96026346; PubMed-756993;
 RA Frazer C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Utterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-F., Dougherty B.A., Bolt K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
 RT "The minimal gene complement of Mycoplasma genitalium.";
 RL Science 270:397-403(1995).
 RV [2].
 RP SEQUENCE OF 721-847 FROM N.A.
 RC STRAIN-ATCC 33530 / G-37;
 RC MEDLINE-94075230; PubMed-8253680;
 RA Peterson S.N., Hu P.-C., Bolt K.F., Hutchison C.A. III,
 RT "A survey of the Mycoplasma genitalium genome by using random
 sequencing.";
 RL J. Bacteriol. 175:7918-7930(1993).
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -1- COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
 CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
 CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
 CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS
 CC IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
 CC MEMBRANE (BY SIMILARITY).
 CC CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC CC
 CC EMBL: U39712; AAC71534.1;
 CC EMBL: U02261; AAD12527.1;
 CC TIGR: MG312;
 CC Cytadherence; Structural protein.

SEQUENCE 1139 AA; 130531 MW; 0011D3288C3DD856 CRC64;

Query Match 3.3%; Score 176.5; DB 1; Length 1139;

Best Local Similarity 18.6%; Pred. No. 0.6; Indels 247; Gaps 39;

Matches 160; Conservative 129; Mismatches 322; Indels 247; Gaps 39;

265 SENTOSLRELKELDPSAQRYSSED-----GLVEDPAKIIISRPN-GVAIPHGDHYFIY 318
 DB 2 ANKQSVFEE-----KNYTOEPENINGDLYDKSVTEEDPNIVAAVDADONGYIYAF 54
 319 SKISAL-----EELIARVPISGTGSTVSTNAPNEVSSLSLSPSSLTTEL 370
 55 NKESTGYVDPYGDTEYDISQLFEDENGPNPFVDEKOEENDYLKTVG---NPDGSDYENG 110
 371 SSSADGTYIPNDIYETATAYVRRGDHFIYIPKSNQIGOPTLPNNLSATPSPSLINP 430
 DB 111 EWWSYTFENDOWI-----STKESQPTDENTYGFDSLP-----P 144
 431 GTSHEKHEEDYGFDPANRI-----IAEDSGF-----VMSHGDHNYFP 469
 DB 145 EVKQPSVEEDNCFD-NOLPREVKQPSVEDNNGFNDLPREVKQPSVVDQSSDYFA 203
 470 KKDITEBOIAPAAOKHLEEVNTHNGDISL-----SHEODYFGNAKEMK 513
 DB 204 KOP-TDENGFNDLPPEVQOPESVVDQSSDDHFAKQPESTYDSFDLPQPTIDOP 262
 514 DLKTEKTEKJAGIMKOYQKRESIVYKKNALIIYHGDHNDPIDEHNPVIGSHSN 573
 DB 263 SLDDHV-----QYNF-----DHR-----EELKPV---AERONN 287
 574 YELFKPEGVAKKKNVYVGEELTVNVLKSTFNQNTFLANGKRVSEFPTELK 633
 DB 288 YOV-----GDOVOAN-LDNNEIOPRAEKVTTDESKAQVVD-----SYQLPIDTQ 336
 634 KLGIMNLVKLIT-PDGKVLKESGVKYGEGVGNANFELDQPLPGGTFYTTASKDYPE 692
 DB 337 QDQTFSSFEQPTVQEQFOVNSEV-----NDQFPEITKEVYLE 377
 693 VSDGTCTVPTSLAYKASOTIFYPFAAGDTYLKVNQFAVPKGTALVYF-----744
 DB 378 SSEFKQDVETS--DINSSENLSNNKDA-----INDSLSEFIOLNSENSE 423
 745 ---DEFHGNALYLE--NNYKVEIKLPIPKLNOGTTTRAGNKIPYFVA---NAVILNOS 795
 DB 424 TASDVAHYEKSESEIHRYKGS-----DLSQSNNSNLSSEPKFSEAPDAHFESQS 477
 796 TYI--VEVPILEKEN---QTDKPSILFOFKRKAQENSKLDER-----VEEPT 839
 DB 478 EPVVOGYDIYQNELKPTLDQSSDDYFAKQPTDENTYGFNDLPPEVKQPSVVDQSS 537
 840 SEKVEKELSTGNSNLSLEVPYVDP-----VQ-----EKVAKFAESYGMKL 884
 DB 538 DDBAKQPESTDSYSDSDLPQ-PTLDQPSLDHVOYNFDHHEELKPVAEONNVOYGF 596
 885 ENVLFNDGTFELPLPGEVYKKNMADFTGEAPQNGENKPSSENGKSTGVENQPTENK 944
 DB 597 DOVQANLDNNEIYQ-PALE--KEYTTDESKAQ-----VYDSYQPLID-DOQOQT 645
 945 PADSLPEAPNPKPYKPNSTNDGNLNPENGNGSDPMLDPALEAPVAPVQEKLEKTAS 1004
 DB 646 FSSSEPOPTVEGPDQVNSEVNDQFKE--ITKEPVLESSEPNKODVVE-----TSN 694
 1005 YGLGDSYTFNMGTIEL 1022
 DB 695 YTNMLQKEDIQSDNKITI 712

RESULT 15
 IGA_NEIGO STANDARD; PRT: 1532 AA.
 AC P09790;
 DT 01-MAR-1989 (Rel. 10, Created)

01-MAR-1989 (Rel. 10, last sequence update)
 DT 01-OCT-2000 (Rel. 40, last annotation update)
 DE IGA-SPECIFIC SERINE ENDOPEPTIDASE PRECURSOR (EC 3.4.21.72) (IGA
 GN IGA.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_Taxid=485)
 RX MEDLINE-90154052; PubMed-2105953;
 RA Bachovchin W.W., Plant A.G., Flentke G.R., Lynch M., Kettner C.A.;
 RT Inhibition of Iga1 proteinases from Neisseria gonorrhoeae and
 RT Hemophilus influenzae by peptide prolyl boronic acids *;
 RL J. Biol. Chem. 265:3738-3743(1990).
 CC - FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.
 CC - CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
 CC CERTAIN PRO-1-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
 CC SUBSTRATES ARE KNOWN.
 CC - SUBCELLULAR LOCATION: SECRETED.
 CC - DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X04835; CNA28538.1;
 DR PIR: A26039; A26038.
 DR MEROPS: S06.001;
 DR InterPro: IPR000710;
 DR PRINTS: PRO0921; IGASERPTASE.
 KW Hydrolase; Serine protease; Zymogen; Autocatalytic cleavage;
 KW Transmembrane, Signal.
 FT SIGNAL 1..27
 FT CHAIN 28..986
 FT PROPEP 987..1532 HELPER PEPTIDE.
 FT ACT SITE 278..278 POTENTIAL.
 FT SITE 986..987 CLEAVAGE (AUTO-).
 FT SITE 1018..1019 CLEAVAGE (AUTO-).
 FT SITE 1121..1122 CLEAVAGE (AUTO-).
 SO SEQUENCE 1532 AA; 168976 MW; 68FF4112BD22F40D CRC64;

Query Match 3.3%; Score 176.5; DB 1; Length 1532;

Best Local Similarity 19.1%; Pred. No. 0.91; Indels 437; Gaps 60;

Matches 247; Conservative 157; Mismatches 451; Indels 437; Gaps 60;

16 VLSL-CAYALNQRS-----QENKNNRVSYYDGSQSSQKSENTLPPQ 58
 DB 11 ISLSIFLAYALTPYSEALVADVDYQIFRFALKNGKFFGAVDLSYKNNRGNI-GNA 69
 59 VSGKEGI-----QAEQYIKITDQGYTS-----HGDIHYNNKRVPPDALFSEEL 105
 DB 70 LSNVPMIDFSVADVKNRIATVVDQYAVSVHAKAEVHTFYGYNGHNDVADENEYRV 129
 106 MKDQNYO-----LKDADI-----VNEVK-----GGYIIRVD-----131

Db 130 VEONNTEPHKMGASNLGLEDYNNANPNFVTEVADIAPIADGGLDITYKDKNRFSSFV 189
QY 132 ---GKYVYVKKDAHAD-----NVRTDEI--- 153
Db 190 RIGGRQLYVEKGYVHOBEGNEKGYDLADLSQAYYALAGIPYKINDQNTNTEGLIGFG 249
QY 154 NRQOEHVKNDEKYSNVAVA-----RSQ-----GRYTTNDGY----- 186
Db 250 NNNROYSAEELKQALSDQDALNTGVLGDSGPLEAFKQKNQWYFGLTYDYMAGYGKSW 309
QY 187 ---VEMP---ADIEDTGNAYIYPHGHYIYIPKSDLSASELAARAHLAGKNMPSOLS 240
Db 310 QEWNIYKEEPADKIKODNAGYVNGEHHM---KTGTNSHIGSTAVRLA--NNEGDLNNG 366
QY 241 YSSASDNNT-----OSVAKS-----TSRPAKSENLQSLKELXDSPSNQRSSES 287
Db 367 QNTYFEBNGLYLVNQNINQAGLFEFRGDTYVKGAN---NDITWLAGID----- 413
QY 288 DGLVDFPAKIIISPTNCPVALPHGDHYHFIPYSKLSALEKTIAR--WPIISCTG----- 338
Db 414 ---VADGRKYVQVKN---PNGDRL-----AKIGKGLEINGTGVNOGLKV 454
QY 339 ---STVSTNAKPNE-----VVSLSLSNPSLSLTSKELSSASDG--YIENPKD 383
Db 455 GDSYVILNQADADKRYOASQYVIGRGTYLVNSNMQINPDNLYFGFRGRLDANGND 514
QY 384 IVEE-----TATAYIVHG--DHFHYIPKSNQIGOPTLPNSLATPPSLPINTGSHK 437
Db 515 LTFEHIRNVEGARIVNHNTHASTI-----TLTGKSLITTNSLSVH--STONDY 563
QY 438 EEDGYC--PDANRIADESGEYVNSH-----GDHNYEPPKDLTE 475
Db 564 DEDDYSTYRPRPIPOGDKLYKNRYALKSGRLNMPENGVAENNDAIFMGYGE 623
QY 476 EOLKAQKHL-----EEVKTSHNGDLSLSHEDYPPNAKEKDLDKITEK 522
Db 624 EAKKNMNNNNRRIDEGFEGDEENGKGNALNL-----NFGKSAQRFL----- 671
QY 523 IAGIMQYVKRESIYVNEKNAIIPPHGDHHDADIDEHKPVGIGSHSNEL----- 576
Db 672 LRG-----GANLNGKISYQGNVLSGRTPHARDFYVKSARAKDAHFSKNEEYVFEEDW 726
QY 577 ---FPEE-----GVAKKEGKV-----YTGEL 597
Db 727 INRTFAAIEIAVQASFSFSSGRVSDITANTATADNAKVNLGKNGDEVCVSDYGYT 786
QY 598 TNYVILLKSTNNQFTLANGOKRVSFSPPELEKLGILNMLVKLITPDGYLEVYSGK 657
Db 787 CNTGN--LSDKALNSFATRNG-----NVMNQNMAVL--GRA--ALMGK 827
QY 658 VFGEVGNIANFELDOPYLPQOTFKYIASKD-----YEPVSYD 696
Db 828 IGGGGSRSYSLNOSHSMHLTGDSOVHNLSDSHILNNASDAQSANKYHTIKINHLGN 887
QY 697 GTFYVPTSLAYMASOTIEYPPHADTYLKNPQFAPV--KGTDL--VRVDEHGNAY 752
Db 888 GHFHYLTLDKMLGKRVLYKESASGHYOLHVONTGEPNQGILDLFDASSVQDRSLFVS 947
QY 753 LENNY--KVEILPIPKLNOGTRT---AGKRIYTFMANAYLNOSTYIYVPILEKE 807
Db 948 LAMHYVDLALNTYI--KTENGITRLYNPYAGNGRPVKPAPSPAANTASQ-----AQKA 999
QY 808 NOTD-----KPSIL-----POFRKNAQENSKLDEKYEETSEKYEKELSETGSTNS 858
Db 1000 TQTDGAQIAKPNQIVVAPPSPQANQAEELRQQAALQAYKROQAALAEKVAKQDEAKR 1059
QY 859 TLEEVPTVDPVOEKYAKPAESTYGMLENVLFNMGTIELYLPSEGVYIKNNADFTGEAPQ 918
Db 1060 KAAETAR---QOEBAKAA-----ELAANKQAE----- 1084
QY 919 GNGEKPSENGKVSCTYENQPTENKPADSLPEAPNEKVPKPENSTD--NGMLNPEGVNG 976
Db 1085 --AEKKAELAKQKAEASHQ--ANAKPRRRRRRRLPRPAPVFSIDDYDAKDNSESSIG 1141

QY 977 SDPULDPAL-----EAPAVDPVOEKLEK 1000
Db 1142 NLAHVIPRQGRLEINDYEELP--LELEDEAE 1172

Search completed: September 26, 2001, 22:34:29
Job time: 233 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:29:21 : Search time 73.11 Seconds
(without alignments)
1880.248 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 5206
Sence: 1 MKRSKTIAGSNAVIVSLSL.....TELALPGEVIAKNLSDFIA 1039

Scoring table: BLOSUM62
Gapop 10.0 / Gapext 0.5

Searched: 425026 segs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPREMBL16:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_unclassified:
13: sp_vertebrate:
14: sp_virus:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	921.5	17.0	822	2	09ZHG7 streptococ
2	230	4.3	2485	5	096134 plasmodiu
3	226.5	4.2	1139	5	097237 plasmodiu
4	223	4.1	1964	2	059947 streptococ
5	218	4.0	1873	2	0924N7 streptococ
6	215	4.0	1694	5	09NXX1 plasmodiu
7	214	4.0	1684	5	09T7T5 plasmodiu
8	214	4.0	1733	2	09K114 staphylococ
9	214	4.0	1939	5	025662 plasmodiu
10	211	3.9	2647	5	09UXX0 plasmodiu
11	210	3.9	1704	5	09T7T4 plasmodiu
12	205.5	3.8	1720	5	025922 plasmodiu
13	198.5	3.7	1653	2	09L8G3 streptococ
14	198	3.7	1698	2	09LCO0 staphylococ
15	196.5	3.6	881	2	09R3J4 streptococ
16	196.5	3.6	1271	5	025860 plasmodiu
17	196.5	3.6	2910	10	09PND5 arabidopsi
18	196	3.6	5005	2	09P8Z5 ureaplasma
19	195.5	3.6	1927	2	054875 streptococ

20	193	3.6	3724	5	077320 plasmodiu
21	191.5	3.5	1236	5	09GTX2 plasmodiu
22	190.5	3.5	4667	5	09TV19 caenorhabd1
23	189.5	3.5	3257	5	09V736 drosophila
24	189	3.5	1826	10	09FM57 arabidopsi
25	187.5	3.5	2151	5	09NG79 trichomonas
26	187	3.5	1080	2	048579 listeria iv
27	187	3.5	4688	2	09P008 ureaplasma
28	186	3.4	1021	2	048581 listeria iv
29	186	3.4	1231	2	P72362 streptococ
30	185.5	3.4	1071	10	09M0D9 arabidopsi
31	185	3.4	2166	5	09V6Y5 drosophila
32	185	3.4	2195	3	002822 saccharomyc
33	184.5	3.4	854	2	09S3P8 streptococ
34	184	3.4	872	2	09S4J3 streptococ
35	184	3.4	1591	11	P97868 mus musculu
36	184	3.4	2394	6	077610 bos taurus
37	184	3.4	3381	6	077609 bos taurus
38	183.5	3.4	2478	2	09LCH2 staphylococ
39	183	3.4	1260	2	09XDB6 streptococ
40	182.5	3.4	1023	10	09LX42 arabidopsi
41	182	3.4	1946	5	097291 plasmodiu
42	181	3.3	1979	11	088196 mus musculu
43	180.5	3.3	2478	2	09RL69 staphylococ
44	180.5	3.3	5170	5	017490 caenorhabd1
45	180.5	3.3	6994	5	017343 caenorhabd1

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	822 AA.
09ZHG7	09ZHG7	01-MAY-1999 (TREMBLrel. 10, Created)		
AC	09ZHG7	01-MAY-1999 (TREMBLrel. 10, Last sequence update)		
DT	01-MAY-1999	(TREMBLrel. 13, Last annotation update)		
DT	01-MAY-2000	(TREMBLrel. 13, Last annotation update)		
DE	HPOTHETICAL 92.4 KDA PROTEIN.			
OS	Streptococcus agalactiae.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_Taxid=1311;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-R268;			
RX	MEDLINE-99115568; PubMed-9916102;			
RA	Spellerberg B., Rozdzinski E., Martin S., Weber-Heymann J.,			
RA	Schultzler N., Luetticken R., Podbielski A.,			
RT	"Lmb, a protein with similarities to the Lrai adhesin family, mediates			
RT	attachment of Streptococcus agalactiae to human laminin."			
RL	Infect. Immun. 67:871-878(1999).			
DR	EMBL; AF062533; AAD13797.1;			
KW	Hypothetical protein.			
SO	SEQUENCE 822 AA; 92385 MW; 80E4EDF313481F98 CRC64;			

Query Match 17.0%; Score 921.5; DB 2; Length 822;
Best local Similarity 26.9%; Pred. No. 1.2e-42;
Matches 267; Conservative 152; Mismatches 305; Indels 267; Gaps 31;

QY	1	MKRSKTIAGSNAVIVSLSLCAALAHNRQENKDNRRVSVYDGSOSOSKS--ENLTPDQ 58
DB	1	MKRTYGTGVAAILLATHIGSYQLGKHHGATKQNLIVYIDSKGKVAAPRTNKTMDQ 60
QY	59	VSQKRGIAEOIYIKITDGYTSHGSHYVNGKVPYDALFESEELMKPNYOLKADAI 118
DB	61	ISAEKGISAEQIYIKITDGYTSHGSHYVNGKVPYDALFESEELMKPNYOLKADAI 120
QY	119	VNEVKGKGIKVDGKYYVYKDAADNVFTKDEINRQKQEHYKD-NEKTVNSVA----- 172
DB	121	INELIDGYVIAKNGNYYVYVYKPSKRNKIRTKQOIAEQVAKTKKAEKGLAQVAHLSKE 180

```

2y 173 -----VARSGRTTNDGYFNPADITEDTGNATYVHGHHYATPKSDLSASLAAK 226
Db 181 EVAAYNEAKROGRYTTDGYIFSPDIIIDLDGADLYVHGHHYATPKSDLSASLAAK 240
2y 227 AMLA---GKNPQGLSTSSASDNTGTSVAKGSTRPAN-----263
Db 241 AYWQKQGRGAPPS--DTRPPAPGRRAPIPDVTPNGQGHDPNGSYHPPAPRPNDAS 298
2y 264 -----KSENLQSLKELEYDSPAQRYSDELVPDAKIIISRTNGVAIPGDH 314
Db 299 ONKHODEFGKTFEELLDQHLRLDKYRHEBGLIPEPQVAKSNAPGVYVPPGDH 358
2y 315 FIPYSKLSALEKTAARVPISGTGVSTNAKPNEVSSLSLSSPSLTYSKLS--371
Db 359 IIPBSOLPSELMELDRLAGCTDNDG-----GSDHSPS---DKETVHTF 402
2y 372 -----SASGYTPNPDIYEETATAYIVHGHHYATPKSDLSASLAAK 414
Db 403 LGHRKAYGKGLDGRPYDTSAYVSKESIHSDKSVTAHGDHPII--GGEQYEL 461
2y 415 PN--NSLATPSPSLPINFQTSHEKEHEDEGYFD---ANRIAEDESGFYVSHGDHNYF 468
Db 462 DEVANVAVAKAGADLVVALDQEGCKEP--FDTKVSRKTYKQCKVGYIMPCKQKDYF 520
2y 469 FKQDITEQIRAAQHLEEVKTSNGLDLSLSEHODYPGNAKEMKDLKTEKTAGINK 528
Db 521 AAYQDLDQIAAEQELMKDKKHRYI-----VDTGLEPRALAVVS 563
2y 529 QYGVRESIYVNEKNAIIPYHGDHHDHDP--IDHKKPVGIGHSHVTEFKPE--SGV 583
Db 564 SUPMAGNATYDTSSEFY--PHIDHIVPYSWLTNRQIAT--KIVQHPVPRDV 617
2y 584 AKKEGNKYTGELINNVNLKNSFTNNQNTLANGOKRVSEFPPELEKLGIMLVKL 643
Db 618 WSKPHE--ESGVIINVPFLDKRAGMNMQI-----HSALEVKALAE--RF 663
2y 644 ITPDGKVIKSVGKVGSGVNIANFELDQYPLPGQTFKYTIASQDYEVSYDGTFTVPT 703
Db 664 AAPDGYIDP-----PRDVLAKETP-----VMDGSGFSIPR 693
2y 704 SIATYMAQOTIFYPPHADDTIARVAPQAVRKGTDALYRVDEFGNAYLENNYVGEIK 763
Db 694 A-----694
2y 764 LPIPLNGTTRTAGNKIPYEMANAYLDNOSTYLEVPILEKEN-----OTDKPSILPO 818
Db 695 -----DSSSRTINKDL-----SQAEWQAOELAKKAGATITDKREEQO 738
2y 819 FERNRAQENSKLDEVEEPTSEKYEKELSETGNSNSTLEEVPTD---PVQEKVA 874
Db 739 --ADKSNEN-----QDPSEASKERE-----SDDFIDSLPQVGLBRATLEHDIN 780
2y 875 KPAESTYGMLENVLENDGTIELYIPSGEVI 905
Db 781 QLAQANIDPKLILIQPG--VOFYKNGELY 810

```

File: us-09-471-255-2.rspt
 Title: 2001 (Trembl) 16: last sequence update
 SER/THR PROTEIN KINASE.

GN PEB0150C.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_Taxid=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 EX MEDLINE=99021743; PubMed=9804551;
 RA Gardner M.J., Tetteijn H., Carucci D.J., Cummings L.M., Aravind L.,

```

RA Koonin E.V., Shallem S., Mason T., Xu K., Fujii C., Pederson J.,  

RA Shen K., Jing J., Aston C., Dai Z., Schwartz D.C., Pertea M.,  

RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  

RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.,  

RT "Chromosome 2, sequence of the human malaria parasite Plasmodium  

RT falciparum".  

RL Science 282:1126-1132(1998).  

CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.  

DR EMBL; AEO01376; AAC71820.1;  

DR HSSP; P24941; 1CKP.  

DR InterPro; IPR000719;  

DR InterPro; IPR002290;  

DR Pfam; PF00069; PKinase; 1.  

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.  

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.  

DR SMART; SM00220; S_TKc; 1.  

KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  

SQ SEQUENCE 2485 AA; 293766 MW; 4297C5CF030AD7E CRC64;

```

Query Match 4.38; Score 230; DB 5; Length 2485;
 Best Local Similarity 18.08; Pred. No. 0.00074;
 Matches 209; Conservative 154; Mismatches 397; Indels 400; Gaps 47;

```

Oy 12 SAVYSLCAIALNQHSEKNNRVSIVDGSOSKSENLTPDOVSQKGIAGIY 71
Db 553 SISLNNVFFYKKRKSNNNNNNNNSSSSSSSKNNHVI--INR-----598
Oy 72 IKIPDQVYTSHGDIHYNGKVPYDALFSEELM-----106
Db 559 -----ISSNIHYKRSKSFENFLEFKILPSKDKCVFERQKDLFEKSEHIK 650
Oy 107 -----KDPYQLDADIVNEVKGIIYKDKYVYLLKDAHADN 146
Db 651 CVSPFNMTSDISSHSSVSKKEPFALKNSIRHLPKENNIITYGSKFNNHQDEKTVL 710
Oy 147 VTKDEIRKQOEHYKNEKYNVAVARSGRTTNDGYFNPADI-----IEPTG 198
Db 711 LKKKEIN--DKNFFSSCLINHNIT--YTLQGVKNKMLGIRDSIYKIDEN 761
Oy 199 NAYIVPHGHYIIPKSDLSASELAARAHLAGKMMQSSLSYSTASDNTQSAVGST 258
Db 762 NMLEECYNGNDSNNKKKKKKKLSFS--CDIINDNITP--YESDKKNSNNKKSMDI 816
Oy 259 SKPANKSENLSLKELEDS--PSAQRKSESDGLVDPAKIISRTNGVAIPGDHNYF 316
Db 817 FNIYKRKSNLNNLSNDSIVDANKINSEYI-----NIQRNKIYELSNKRIINYK 870
Oy 317 FYS-----KLSALEKTAARVPI-----334
Db 871 LYSDEIRKYSLEKKEYIDNISNNRERYTKNEMINERISKMDILYPCDRKNSLMSGP 930
Oy 335 -----SGIGSYSTAKPNE-----VYSSIGSLSSNFS 363
Db 931 VTIENNISRENEKNSVILKKKKNENFCVGLCHGKKNNDNITIDGNIKKNEE 990
Oy 364 LITSKELSSASDGYIFNKDIYEETATAYIVHGHHYATPKSDLSASLAAK 423
Db 991 ITKDEYISREKKNYK-----CIRNPDIKY--EYV-----1022
Oy 424 PSLPINFQTSHEKEHEDEGYDANRIADE-----SGFVSHGDHNYFFKQD 472
Db 1023 -----LSVHTLDEDKKKDNNMLIDNNNEALIEYVNGVYNNIILDRKNNNS--RKD 1071
Oy 473 LTEQIRAAQHLEE--VKTSHNGDLSLSEHODYPGNAKEMK--DLDKIKEKTAGINKO 529
Db 1072 MEKEMEKEMEKEMEKEMEKEMEKEMEKEMEKEMEKEMEKEMEKEMEKEMEKEMEK 1126
Oy 530 -----YGVKRESIYVNEKNAIIPYHGDHHDHDPIDEH-----PVGIGHSHSNY 574
Db 1127 EINTYKNNIYVDKLEIYNEKK--LIYFPNYS-----DYKKNMMSININCDDY 1180
Oy 575 -----ELPKPEGVAKKE-----GNKYYT--GEELTVNVNLK 605

```

[illegible]

QY	150	KDEINRKOGEVYKZENEY-----NSNVAVABSG-QRYTN-----DGYT	187
Db	154	KNETNDLENEVYVSMOVSYEVLNDNFISRSTSIYNDJNSEVLRKENIKSSHHNGY	213
QY	158	FN-----PADIEPTGNAVYVPHGHHYVFKSDLSASLEAAFAHLAQ--NMOPOLS	240
Db	214	INDHNIKSDIINDHNI-----KSDI--INDYDNIKGGIKNEYINDIKSDKN	259
QY	241	-YSSYASD--NNTOSVAKGSTSRPANKSENLQSLKELYDSPAQYSSSDLVDPAKI	297
Db	260	DYDNKISDIIINDYDNI-----KSDKINDYDNKSDKINDYDNKSDKINDYDNKISD--KI	313
QY	298	ISRPNGVAVPHGHHYVFKSKISALE-----EKIARMPISGTSYVSNAPREYVS	353
Db	314	-NDYDNKISDKINHYDNKISDKINDYDNKISDKINDYDNKISDKINDYDNKINDD--	368
QY	354	LGSLSSNSSLTTSKELSSASDGYIFMPKOIV-----EETATAVYVRHGDHPHYIPKSNQ	408
Db	369	-SQJNNVSNKVNSPD--DBDNYTEKDKIILINCNIYEEN-----KNMA	409
QY	409	IGOPTLPNNISLATSPSL-PINPQTSHEKEEDGYPD-----ANRIADESGFYWASHD	463
Db	410	-----NNIHNKENIHMYDGNKNEENNNDINKNNDINKVNLTRVDTEYTLNEEN	462
QY	464	HNHYFFKDLTE--EQIAKQAKHLEEKYSYHNGDLSHEDODYPCNAKEMKODLKIE	521
Db	463	ILPLTNDYDVKLDEINDKNTIENKKEK--INLSKFAVTLNKKKELEBELKICEN	520
QY	522	KIAGIMQYGVKRSIYVNNKNAIIPHDHHDAPIDBHKRVGSHSWELFKPE	581
Db	521	KEEQI-KMY-----NKNLQNKDEIYANVNIQNKDKIESY-----HIKLNKEE	565
QY	582	GVAKREGKRVYTGELTNVNLKSNFNNQ-----NETLANQKRVSEFSPELEK	634
Db	566	EIANVNIQKKEEIIKYVNELONSVOYKKEKELKTINVLKNTMEKLE-----KEISL	621
QY	635	LGIMNLVKLITPDGKVL-----EVSQGVFBGSGVIANFELDQYLPDQYFK	682
Db	622	LG---EMKIEEENKVLKIFKEEKNKINNKEV-QIGGDINDNDMKFKQILLNKFTN	677
QY	663	YTIAS-----KQPEY-----SYDGTFFYPSL-----AYKASQT	713
Db	678	KLVNELNIYVENILFCFSIDKKNRLLVCIKONDNFVIPCIDEKFPGLQKTESM	737
QY	714	IFYPFBADTV-----LRVNPQFAPKGTDLAVRDEFH--GNATLENNY--KVGELK	763
Db	738	IIODIHEEELLNKQONELIYIKKKEKDIYEIEBKAKYATSLNEINNYKNKEE	797
QY	764	LPPLKNGTIRTAGNKIPVYFANAALDQSTY-----IYVPILEENQTD	811
Db	798	NEIKELNN-----TLNKYKEEMNPKREIYIYNEKYALIEIJECKERIND	843
QY	812	KPST-LPOFKRKAQENSKLDEKVEEYKISEYKEREKLSGTGNSYNSLVEVPYVDPQ	870
Db	844	QOANYGSIOLKFKMLKEKIELEKKYKE--QYEKETOORINDMKIIFDNKKEIIOQIDIDL	901
QY	871	EKVAKKAEYSYAK--LENVLFNNDGTIEL-----YL	899
Db	902	HKIEKLFSSNDEKRTIENQIYDNDNDNIMQEGIQKNEKKQIIDNPINTDEKOLLNTI	961
QY	900	PSGEV-----IKRMAADTGEAPQNGENKRPENGVSTG	934
Db	962	PDNHNISDLNHNINVENDYVEKRNKNVHLSKKEINYOQHINIEGNDONNNNNNNNN	1021
QY	935	TVENQPTENKPADSLPAPNE-KVPKRENSDNGALNIPGANGSDPMLDPALEAPVDP	993
Db	1022	NWEKKSINSYNTSIPYIPEYKIRKIKKETEYIELLN-EOGBGKKKM--TEEINSIKN	1076
QY	994	VOEKLEKFTASYGIGLSDVYFNNDDGIELRLPBGVEYIKKNLSDETA	1039
Db	1077	QVKNYVESINGNY--OHIIYQ-----KNILSNFTA	1103

QY	704	SLAVKASQITIFPFPAGDQITLVNQ	----	FAVPGTALVRVDEFHGAYLENNYK	759
Db	444	AVV-----	TQGEI	-EYQESPDYVSDKGEQYAPLPEYKGN-----	I 482
QY	760	GEIK-LPIPKL-NOGTTTAKRIIVPTMANVLDNOSTYIVVPILEKENDKPSIL	816		
Db	483	EQVAVETPVEVTEKQGPBEKT--EEVYV-----	KPIETPVVNDNEGTVGSGI-	527	
QY	817	POFARN---KAQENSKLDEVEEPEKISEK	-----	EKEKISEGNGSTS	856
Db	528	-QEAENVPQAEESTTJNSEKVSPTSISENTGEVSNPSDSTTSVGBESKNPEINDSKNNS	586:		
QY	857	NSTLEEVPTDPOAEKAVKFAESYGAKLEIVLFINDDGTEILPLSGEYKKNMADGTGEA	916		
Db	587	EKIVEEVP-VNP-----	NGGITE-----	GTSNQETEK	612
QY	917	PQNGENKRPBENGKV---STGVENOPTENKP--ADSLPEADNEKVPKENS--	-TNGS	967	
Db	613	PVQPAEETQJNSKIANENTGEVSNRPSDKPVEESNQPEKNGTATKPENSGNTTSENG	672		
QY	968	MLNPE	972		
Db	673	QTEPE	677		
RESULT	5				
Q924N7	Q924N7	PRELIMINARY;	PTI: 1873 AA.		
AC	Q924N7				
DT	01-MAY-1999	(TREMBLrel. 10, Created)			
DT	01-MAY-1999	(TREMBLrel. 10, last sequence update)			
DT	01-MAY-2000	(TREMBLrel. 13, last annotation update)			
DE	SURFACE PROTEIN PRECURSOR.				
OS	Enterococcus faecalis (Streptococcus faecalis).				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;				
OC	Enterococcus.				

Accession	Strain	Host	Location	Year	Genome	Genome Size (Mb)	GC Content (%)	Genome Type
RP	SEQUENCE FROM N.A.							
RN	[1]							
RC	STRAIN-MH594;							
RX	MEDLINE-99081742; PubMed-9864215;							
RA	Shankar V., Baidyan A.S., Huyke M.M., Lindahl G., Gilmore M.S.;							
RT	"Infection-derived <i>Enterococcus faecalis</i> strains are enriched in esp,							
RT	a gene encoding a novel surface protein.";							
RL	Infect. Immun. 67:193-200(1999).							
DR	EMBL; AF034779; AAD09838.1; ..							
DR	Interpro; IPR001899; ..							
DR	Pfam; PF00746; Gram_pos_anchor; 1.							
KM	Signal.							
FT	SIGNAL							
FT	1	49	POTENTIAL.					
FT	50	1873	SURFACE PROTEIN.					
SO	SEQUENCE	1873 AA; 202084 MM; F609483DB980621 CRC64;						

[illegible]

```

Db 739 TVSKN-----IYENGENIPAGYHKVFTAGEGTSIESGTTVEAVKGVSLPE----- 786
QY 315 FIPYKSLALEKIRAMPISGTSVSTINAK-PHEVYSSLSGSSINPSLSLSTSELSA 373
Db 787 -----DKLPYKAKKDG-----YTDAMPE-----ATQPKADDFEVSSA 822
QY 374 S--DGYTPKRC-----IYETATAYVRGHDHRYIPKSNQIGOPTL 414
Db 823 TKLDDIIEPNPAGIHKVFTAGEGTSIESGTTVAVKDG-----VSL 868
QY 415 PNNSLATPSPSLPINTGSHHEKEDGCGFDANKRIIADESGFVNS-----HGDD 464
Db 869 PEDKL---PYKAKDGYDAMPE-----EATQPKADDFEVSSATKLDDIIEPNGN 919
QY 465 -----NHFFPKD---LTEBQ---IAAQKHL----- 486
Db 920 IPAGYHKVFTAGEGTSIESGTTVAVKDGVSLPEDKLPLYKAKDGYTDAMKPGEATQPI 979
QY 487 -----EVTSHNGDLSLSEHODYPQNAKEMDLKIEKLAGIKQYGVKRESTITVNR 541
Db 980 KADDFEVSSATKLDDKSDADKYTEPGOKVYTELAKKE-PDASEGIKKNDLKRDAKYTK 1038
QY 542 E-----KNATYPRGHDHMDPIDHKPVGIGSHSNYTELKEPE-EGVAKKEGK 590
Db 1039 EKVDTISAGNKKGTVVYTS-D-GSSDEVE--VDVYTDNRSDADKTEPVEGEKEVIGK 1095
QY 591 VTTGSELIVNVLKNSNTNN-----ONFTLANGQKAVSSPPELEKIKGINLVYL 643
Db 1096 VDLTDVNTMLPPLPGQTYTDTVTPGGTIDTNTPGNEVIEVTPD-----GTRDVTAV 1149
QY 644 ---IPPDGVLEKVSQKVGEGV-----GNIANFELDQPYLPQGTGKTYTASKDP 691
Db 1150 PVEVTDNRSDADKTEPVEGEKEVIGKVDLDNVTNL-----PTLPQGT---TVT----- 1197
QY 692 EVSVDGTFTVPTSLAKMASQTIFFPHAGDLYLRVNPQFAVPKGTDALVRVDEEHGNA 751
Db 1198 DVTGGTIDTNTPGNEVIEV-----TY-----PDGKDVAVKVPVEVTDNR 1239
QY 752 YLENNY-----KV-----GEIKLP-----IPKLNQGT---RTAGKIKIYVTRMAAYILDN 793
Db 1240 SDADKTEPVEGEKEVIGKVDLDNVTNLPLPQGTYYTDTVTPGGTIDTNTPGNEVIEV 1299
QY 794 GQTY-----IYEVPLKEKNOTDKPSILPQFRNKAQSNKLDKEVEBPKSEKYEK 847
Db 1300 EYVTPDGTDTYKVPVEVTDNRSDADKTYPMVEGEKEVIGKVD----- 1343
QY 848 LSTGSGTSTNSTLEEVPTVDYQVEKVAKEASGKMLENVLENNQGTIELYLPSSG---EVI 905
Db 1344 --LTDVNTMLPPLPGQTYTDTVTPG-----GTIDTNTPGNEVIEVTPDGTGKDV 1393
QY 906 K-----KNMADFTGAPGNGENKPSSENGKYS--IGTVENQPLENKPADSLPEAPEK 958
Db 1394 KVPVEVTDNRSDADKTEPVEGE--KVEIGKVDLDNVTNLPL-----LPQGTVTV 1445
QY 959 KPENSTDNGLNLEGN 974
Db 1446 TPEGCTIDT---NTPGN 1458

```

```

RESULT 6
Q9NHX1 PRELIMINARY; PRT: 1694 AA.
AC Q9NHX1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE MAJOR MERZOLOITE SURFACE ANTIGEN.
GN GP195.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-FCC1/HN;
RA Shan 2.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT "Molecular cloning and sequence analysis of major merozoite surface
RL submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218248; AAF27526.1;
DR InterPro; IPR000561;
DR InterPro; IPR003247;
DR Pfam; PF00008; EGF; 1.
DR ProDom; PD001527; 1.
KW Merzoite.
SQ SEQUENCE 1694 AA; 192766 MW; B51634A49E0F6728 CRC64;

Query Match 4.0%; Score 215; DB 5; Length 1694;
Best Local Similarity 20.3%; Pred. No. 0.0027;
Matches 248; Conservative 181; Mismatches 409; Indels 384; Gaps 67;

QY 1 MFSKKYIAAGSAVIVSLCAVALNQHRSQENKDNRRVSYDGS---QSSQKSENTLP 56
Db 500 MFEFNNF---DKDVYDKIFSAITYN---VEKQYNNKRSNNSTYVYQKLLKALSTYLE 553
QY 57 DOVSQKEGIAQIYIKITDQGYTSHGDHITYNGKPYDA--LFSREL----- 104
Db 554 D-YSLRKGJ-----SEKDPNHYTLKTGLLEADIKRLKEIKSSENKJLE 596
QY 105 -----LMDPNYQLDADIVN-EVAGGYII-KVGD--KYVYLKDAHADNVRT----- 149
Db 597 KNEKGIHSAASLESVDVVKQVAVLTIKTIEDIKLELKLKNAQLKDSIHVRNIYTP 656
QY 150 -----KDEINRQKQ-----EHVKDNEKYNVA---VARSQGRYTDNG-- 185
Db 657 QNKPEPYIYLVKKEVDKAEIIPYKMDLKEQAVLSITQPLVAAS--TTEGSGHS 713
QY 186 -YVNPADIIEDTGNAIYVHGGHYHI-----PRSDLSASELAAKAHLAKNNQPSOL 239
Db 714 THTLSQSGTEETEETV---GHYTVYITLPTQPSPEKVEKVENSIERK----- 761
QY 240 SYSSASDNNTQSVAKGSKRPAN-----SENLSLKELYD-SPSAQR 283
Db 762 -----SNDSQALTKTYTLKILDEFLNYSYCHKIYLVSSNMQKLELVNLRPEKN 815
QY 284 YESDQVLDVDPKAIISRTNGVALIPHG-----DHYHFIPYSK---LSALEEK-- 327
Db 816 ELAKS---CDPLDLEFNQNNIPAKSYLSDSMNNDLQHLFEELYQEMLYLHLKKEEH 871
QY 328 ---IARVNPISGTSYST-NAKPREVYSSLSGSSNPSLTSKEISSADGTYFNK 382
Db 872 IKKLEEQKQITGTSSTSPGNTVTVAQSATHSNQNOOSNASS---TNTONG-- 922
QY 383 DIVERATAYIVRHQDHFYIPKSNQI--GQPLPN--NSLAPSPSLPINTGSHKEHE 439
Db 923 -VAVSSGPAVVEESHDLPTVLISINDLKIVSLNLGNKTYKPNP-LTIST-TEMEKFE 979
QY 440 DGYGFDANKRIIADESGFVNSGHDHNYFFKRD-----LTEQIKAAQKHLDEVTS-- 491
Db 980 N-----ILANNQTYF---NDQIKQFVKNKSNKYITGLTFQKNAALNDEIKKLDTQ 1027
QY 492 ---HN-----GLDLSSEHODYPQNAKEMD---LDKIEEKIAGIKQYGVKKE-SIV 538
Db 1028 LSPFDLYNKYKLLDLRFNKKKELGODKQIKKLTLLKEQLEKLSINPHTVNLQFVSF 1087
QY 539 VKEKNALY-----YPRGDHMDADIDEKPYGI--GSHSNY 574
Db 1088 FKKKEAEIAETENTLENTKILKHYKGLVKKYNGE---SSPLKLTSEVSIQTEENTYAVL 1144
QY 575 ELFKPEEGYAKKEGNVYNGEE-----LTNVVNLKNSFNQNFNTLANGQKRVSS 626
Db 1145 EKFRYLSKIDGKLNDLHGKRLSFLSSGLHLITELK-EVYKNNNTY-----GN 1194
QY 627 PPELEKKIKGINLVYL---IPPDGKVL-----EKVSKVGEVGNIANFELDQPYL 676
Db 1195 SPSENNK--VNEALKSYNPLPEAKVTVTVVPPQDVTPLSPSLRVSSSSSTSENOI 1252

```

677 PGO-----TFRYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPFHAGDTYL-RVNPQ 730
 Db 1253 PFGSLTLELOQVOVOSNDDEED-DSLVLPL-----IFGESEDNDEYLDQVVTG 1300
 QY 731 FAVPGGTALAVRVEDEFGNATLENNYKVEIKLPIPLKNOGTTRTAGKIP---VTFMA 787
 Db 1301 EAVSYTMNLTSGF-----ENEYDIYTLK-PL---AGVYRSLKQLEKNTITFNL 1346
 QY 788 NA-----YLD-----NOSTYIYVPILEKNOGTDKPSIIPQFRN 822
 Db 1347 NLNDLNSRLKRRKFTFLDYLEDLMQFKHISNEYIIE-DSFKLLNSBOKNTLLSKYKI 1405
 QY 823 K-----AOENSKLDEKY-----EKPTESEYKEREKLTSTGNSNSTLEEVPTVP 868
 Db 1406 KESVENDIKFQEGISYEYKYLAKYKDDLESIKKYIKE---EKEPSSPPTTPSPAKTD 1462
 QY 869 VOEKAKFA-----ESYGMKLENYLFNM-----891
 Db 1463 BOKKSKFLPPLTNIETLYNMLVNNIDYLLMLAKINDCNVEKDEAHYKTKLSDLKAI 1522
 QY 892 DGTIELYLPSEGVKKNMADFTGEAPONGENKPSENGK-VSTGTVENOPT-----EN 943
 Db 1523 DDKIDLF-----KNTNDEFAIKKLLINDTKMDLKGKLSLSTGLVQNPPTIISKLEIG 1574
 944 KPADSLPEAPNE--KPVKPENS 963
 1575 KFOQMLNISQHCQVKKOCPENS 1596
 RESULT 7
 ID 097275 PRELIMINARY; PRT; 1694 AA.
 AC 097275;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE MEMOZOITE SURFACE PROTEIN 1 PRECURSOR.
 3N MSP1.
 3S Plasmodium falciparum.
 3C Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 3X NCBI_TaxID=5633;
 3N NCBL_TaxID=5633;
 3N SEQUENCE FROM N.A.
 3R STRAIN-HN1;
 3L Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
 3L Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
 3R EMBL, AF062348; AAC72884.1;
 3R InterPro; IPR000561;
 3R Pfam; PF00008; EGF_1.
 3W Merozoite.
 3Q SEQUENCE 1694 AA; 192794 MW; 84CFCE0E709F5673B CRC64;
 ery Match 4.08; Score 214; DB 5; Length 1694;
 at Local Similarity 20.38; Pred. No. 0.0031;
 Matches 248; Conservative 179; Mismatches 411; Indels 384; Gaps 67;
 Y 1 MKFSKTYIAASAVYVSLICAYALNHRGSEN---KDNRRVSYVDSOSSQSSSENITP 56
 Db 500 MKNFNNF---DNDVYDKTFSAKYTYNEKQYNNKPFSSSNSSVNV---QKLKALSTLE 553
 Y 57 DOVSQKEGIAQEQIYIKITDOGYVTSHGDRHYHNGKVPYDA---LFSEEL-----104
 Db 554 D-YSLARKGI-----SEKDFNHYYTLIKGLLEADIKKLTLEELKSEENKILE 596
 Y 105 -----LMDPNTQLADADIVN-EYKGGITI-KVDG--KTYVYLLDAADAVNR-----149
 Db 597 KNRGGLTHSANASAEVSDIVLYQVQVLLIKKIEIDLRKIEFLKNAQKDSIHVPNIYKP 656
 Y 150 -----KDEINROKO-----EHVKNDEKVNNSVA---VABSGRYTNDG---185
 Db 657 QNRKEPYLLYLKKEVNDLKEEIPRYKMDLKKKQAVLSSTIOPPLVASE---TTEDGSHS 713

QY 186 -YVENPADIIEDTCNAYIVPGCHYHI-----PKSDLSASELAATAKHLAKNNQPSQL 239
 Db 714 THTLQSGSETEETETV-----GHITTYTILPQPSPPKVKNVENSIEHK-----761
 QY 240 SYSSSTASNNFTOSVAKGSTSKPANK-----SENQSLLEKELD-SPSQOR 283
 Db 762 -----SDNSQALTKRYTLKLDLEFLKSYCHKYLIVSNSMQKILEVYNLPPEKN 815
 QY 284 YSESDGLVFPDAKISRPNGVAIPHG-----DHYHFIYKSK-----LSALEEK--327
 Db 816 ELKS-----CDPLDLEFNQNNIPANYSLYDSKNNDLQHLPELYOKEMYYLHKLEENH 871
 QY 328 ----IARVPLSGTGYST-NAKPNFYVSSLSGSLSSPSSLTTSKELSSADGYTFNPK 382
 Db 872 IKKLEEGQKQITGSTSTSPGNTVNTNQSATHSNQOQNAS--TNNQNG-----922
 QY 383 DIVEETAPAYIVRGGDHFHYIPKSNQI-GQPTLPN--NSLATPSPSLPINTGSHKHEE 439
 Db 923 -VAVSSGPAVVEESHDPPLTVLSISNDLGIYSLNMGKTKVPPN-LTIST-TEMEKYE 979
 QY 440 DGYGDARIIAEDSGVYMGSHGDNHYFEKKD-----LREQIKAKQKLEEVKTS--491
 Db 980 N-----ILKNDTYF-----NDDIQFVKSNSKVITGLTETQNALNDEKIKKIDTLQ 1027
 QY 492 -----HN-----GLDSLSEHODYPGNAKEMD---LDRKIEKINGIKQGVKRE-SIV 538
 Db 1028 LSFQLYNKYKLLKDLFKFKKELGODKMQIKKTLTKLEQLESKLSLNNPNNVLONFVIF 1087
 QY 539 VNKERNKI-----YPRGDHHPADIDEKPVGI--GHSHSNY 574
 Db 1088 FNKKKEAIEAETENTLENTKILLKHKGVLVYKNGE---SSPLKLTSEVSJOTEDENYANL 1144
 QY 575 ELFKPEEGVAKREGKVTGGE-----LTNNVNLKNSFNNOANTLNGKQKRVFS 626
 Db 1145 EKFRYLSKIDGLNDLNLKGLKKSFLSSGHLHLTEIK-EYIKANKNT-----GN 1194
 QY 627 FPEPEKRLGIMLVKL---YTPDGKVL-----EKYSGRYFGGVNINAFELDQVYL 676
 Db 1195 SPSENNKK--VNEALKSYENPLPEAKVTVVTPQPDVTPSPLSVRYVSGSSGSTHEEQI 1252
 QY 677 PGO-----TFRYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPFHAGDTYL-RVNPQ 730
 Db 1253 PFGSLTLELOQVOVOSNDDEED-DSLVLPL-----IFGESEDNDEYLDQVVTG 1300
 QY 731 FAVPGGTALAVRVEDEFGNATLENNYKVEIKLPIPLKNOGTTRTAGKIP---VTFMA 787
 Db 1301 EAVSYTMNLTSGF-----ENEYDIYTLK-PL---AGVYRSLKQLEKNTITFNL 1346
 QY 788 NA-----YLD-----NOSTYIYVPILEKNOGTDKPSIIPQFRN 822
 Db 1347 NLNDLNSRLKRRKFTFLDYLEDLMQFKHISNEYIIE-DSFKLLNSBOKNTLLSKYKI 1405
 QY 823 K-----AOENSKLDEKY-----EKPTESEYKEREKLTSTGNSNSTLEEVPTVP 868
 Db 1406 KESVENDIKFQEGISYEYKYLAKYKDDLESIKKYIKE---EKEPSSPPTTPSPAKTD 1462
 QY 869 VOEKAKFA-----ESYGMKLENYLFNM-----891
 Db 1463 BOKKSKFLPPLTNIETLYNMLVNNIDYLLMLAKINDCNVEKDEAHYKTKLSDLKAI 1522
 QY 892 DGTIELYLPSEGVKKNMADFTGEAPONGENKPSENGK-VSTGTVENOPT-----EN 943
 Db 1523 DDKIDLF-----KNTNDEFAIKKLLINDTKMDLKGKLSLSTGLVQNPPTIISKLEIG 1574
 944 KPADSLPEAPNE--KPVKPENS 963
 1575 KFOQMLNISQHCQVKKOCPENS 1596
 RESULT 8
 ID 09K114 PRELIMINARY; PRT; 1733 AA.

AC 09K114
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
 DE 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
 DE PUTATIVE CELL-SURFACE ADHESIN SDRF
 GN SDRF
 OC Staphylococcus epidermidis
 OC Bacteria; Firmicutes; Bacillus/Clostridium group
 OC Bacillus/Staphylococcus group; Staphylococcus
 OC NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9491
 RC MEDLINE=20340957; PubMed=10878118;
 RC McCrea R.W., Hartford O., Davis S., Ni Eldin D., Iina G.,
 RC Speziale P., Foster T.J., Hook M.,
 RC The serine-aspartate repeat (Sdr) protein family in Staphylococcus
 RT epidermidis.
 RL Microbiology 146:1535-1546(2000).
 DR EMBL AF245041; AF272509.1;
 DR InterPro: IPR001899;
 DR PROSITE: PS00343; GRAK_POS_ANCHORING; UNKNOWN_1.
 SQ SEQUENCE 1733 AA; 184720 MW; D8D62EA1692PD4E8 CRC64;

Query Match 4.0%; Score 214; DB 2; Length 1733;
 Best Local Similarity 18.5%; Pred. No. 0.0032;
 Matches 226; Conservative 175; Mismatches 458; Indels 360; Gaps 55;

2 KFS-KYIANGSAYIVLSICAVLNCH
 22 KYSIKRTVGTASTIVATITMGADNEAKAEDNQLSASKEQKSDNENSKLNQVD 81
 41 YVDSQSSQSENL-TPDOVSQKEGIAEQIVIKITDQGYTSHGDHYHYNGKVPYDAL 99
 82 LDNGSHSEKETTNNATVEVKVEAPPTSDVSKRKAENAVTN----- 124
 100 FSEELKKDPNTQKADIVNEVKGITIIYVDGKYTYLLDAHADNVKTRKDELNRK-- 157
 125 -----ESTKPKTEAPTYNEES-----IAETPKSTTQDDSTEKNPRLKDLNLSSTT 173
 158 QEHYKDNENSNVAVARSGRTTNDGYFNPAADIIEDGNAYIVHGHYHPIKSDL 217
 174 SKESKTEHSTKAQKSNSTNSNLDNTD-----SPQSEKTSQA-----NDS 216
 218 SASELAAKAHLAGK-----NMQPSQ-----LSYSTASDNNTQSVAK 255
 217 TDNSASPK-QLOSKPSEQVKTKFNFDEPTQVDEHTTKLTPSVSTDSVNDKQDYTR 275
 256 GSTSK---PANKSENLQSLKELYDSPAQRYSBGLVFDPAKTIISRTPN-GVAIP-- 308
 276 SAVASLGVDSNETEALITNAVDNIDPLAASREQINEAIIAELKKDSNPNDGYDPTLAL 335
 309 -----HGDIYHPIYKSLALEEKIARMPISG-----TGSTVSTNAKNEVYSS 353
 336 NRSGSKSPKASPRNKLASLAE-----PNSGKNVNDKVKITNPILSLKSNHANNV 390
 354 LGSLSNPSSILTSKEL-SASADQYIPNR----- 382
 391 IMPTSNQPIKANKYELDDSIKEDGFTIKYGYIRPGLELPAIKQQLSKDSQSIYANG 450
 383 --DIVEETAAIYRBDHRYIPKSNQIQPILPNNSLTPSPSLPINTGSHKEED 440
 451 VYDVTITTYTTFYTNVDOQNTIGSFDL-----IATPKRETAIK-----DNQ 493
 441 GYGRD---ARIIAEDSGGVMSGHNDHNFKKDLDEEDIKAAQKLEVEKTSHNGDS 497
 494 NTPREVTIANEVAKD--FIVDGN-----KQDNT--TTAAVAVDNNKHNKHNVY 541
 498 LSHEDD-----YPG-----NAKEKDKDKTEE 521
 542 LNNNNPKAKAYSTYKNGEPIGEVKYEVYTDNANVDSFNDLSSNVKQDVTSGAP 601

QY 522 KIGIKQYGVKRESIYVNEKNNAIIPHGHHADPIDEKKFVGIGHSNTEKPEE 581
 DB 602 KYSADTRDINARSMANKKIY-----TQVYRTGNGVYTEWLR--D 647
 QY 582 GYAKKGNKRYTGEELTNVNLKNSTPNQNTLANQ-----KVSFSPPELEKKGI 637
 DB 648 GTTNT--NDEYRGTSKSTYYLNGSSTAQGNPTYSIGDVTMLDKKNNGVODDEKGLA- 704
 QY 638 NMLVLTTPDGKYLEKVSQVPEE-GVGNLAN-----FELDQPLPGQFTKTKSKOV 690
 DB 705 GYVYTLKDSNNRELQRTYDQSGHYQFDNLQNGTYVEFAIPDNTYPPSPANNSTDAID- 763
 QY 691 PEVSYDGTFTVPTSL-----AYKASQITFY-----PEHAGDTYLRNVPQFVPGKTALVR 742
 DB 764 SDEBDGTNRVYVAKGTINNADNMTDGTFLPKTNVGD-----Y 804
 QY 743 VFDEFGNATLENNYK-VGEIKLPILPILNO--GTRTAGN-KIPVFMANAYLDNSTY 797
 DB 805 VMEDTINKDGIQDDNENGISGVKTYLKNKNGDTIGTTTDSNGKYEFTGLN-----GDY 858
 QY 798 IVEVPILEKENQDKPSILPQFRN-----KQENKLDKVEEPTS-----EK 842
 DB 859 TIEFETPEGTPTKONGSDEGRDSNGTKTIVTKADNKTIDSGFYKPTYNAGDYWED 918
 QY 843 VEREKL---SETGNSTNSYLEEPTVDPVQEKVAKFAESYK-NENYFNNDGTIELY 898
 DB 919 TNDGDIQDDSEKISGVKTYLKD-KNGNAIGTTTDSAGHYQKGLENGSY-----IYEF 973
 QY 899 LPSEGYIKKMAADFTGAPQNGENKPSENGKYSTGTVENOPTENKPADSLPEAPNEKVP 958
 DB 974 TPGS-----YTPKANSQDITVDSNGITTTGILNG-----ADNL----- 1008
 QY 959 KPNSTDNQML-NPEGVSGDPMIDPLAEAPVADPVQELKFTASYGIGDSVIFNMD 1017
 DB 1009 -----TIDSGFYKTPKPSVSG-----DYWEDTNK-DGIQDNER-----GISGV 1046
 QY 1018 GTIELRPSGEVYKKNLSD 1036
 DB 1047 -KYLNDKGNITITTTD 1064

RESULT 9
 Q25662
 ID Q25662 PRELIMINARY; PRT; 1939 AA.
 AC Q25662;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
 DT 01-AUG-1998 (T-EMBLrel. 07, Last annotation update)
 DE REPEAT ORGANELLAR PROTEIN.
 OC Plasmodium chabaudi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5825;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=96V;
 RC Weiner E.B., Taylor W.R., Holder A.A.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U43145; AAC63403.1;
 SQ SEQUENCE 1939 AA; 229001 MW; B36E462001CGF22F CRC64;

Query Match 4.0%; Score 214; DB 5; Length 1939;
 Best Local Similarity 17.9%; Pred. No. 0.0038;
 Matches 220; Conservative 183; Mismatches 415; Indels 408; Gaps 53;

28 HROENDNRRVYVDSQSSQSENLTPDOVSQKEGIAEQIVIKITDQGYTSHGDHY 87
 DB 462 HKELNELNQLSKLNKKNINEN-----TELNDKI 494
 QY 88 HYNGKVPYDALFSEELMDPNYQLKA---DIVNEVKGITIIYVDGKYTYLLDAHAH 143
 DB 495 SSLSNEY-----NILNDKQTLGNDITLNDLNNLK----- 526

```

144 ADVKTD-EINROGAEHVKDNKYNVAARSGRTTNDGYF-----NPADIT 194
527 -NEIWTSDKMKKMEADLAMEEKEGCVIIDELEKYNKIFLELEKKEKYNADLN 585
195 EDTG---NAYTPHGGHNYIPKSDLSASLAAKAHLAKGMOQPSOLYSTAD----- 247
566 DEISLIRNITYKE-----KEPIEKEPEYENKINFNKKEKKNILYENELNLSLK 637
248 -NNTOSVAKGSTSKPANKSENLOSLLKEY-----DSPAQRSESDGYF 292
638 YDNESGLKQIDELMIOKLTKEKYL-QIYDNMMMPRSICKIDMPSINIKSGDLVDF 696
293 DPAKIIISIPNGV-AIPGDHYHPIYSK-----LSALEKRIAMVPISGTG----- 338
697 VYAYIKRDESSDANPOTHEKEMVALEKRAALVYALEEKKREIKALGCHGVYLR 756
339 -----STVSINAPREYVSLG-----SLSSNPSLSTTS----- 367
757 LGEOKKEBTIILEEKHKDVYTKLGEOKHENTIKLEEKHDVYTKLGDYKKEIRAKLEEH 816
368 -----KELASASGYIFNPDIYEETATAYVRGDFHPIPSNOIGOPTLN----- 416
817 AVYVALEKKEKLGSGH---KEMVDELEK---NHADPVEGLEKHKETAKLEEGHKS 868
417 -----NSLAPPSPLPIPIGTSH-----EKHEEDGYFDMRIIAEDS 455
869 EKNNEKHNADVELEKHKAEFTAKLGEHREYVAGLEEKKE-----VYALEE 919
456 GFVMSGHGHNHYFFKDLTERQIKR--AKHLEEKYTSHNGDLSISHBODYPNNAKEMK 513
920 -----HKKEIALKEGHEKEMVALEKKEHVA---GLEAKHNLEEGKEKEMVALE 968
514 -----DLDKKIEEK-----IAGIMKOYGVKRSIYVKNKEMITPHGHH 554
969 KKHADLVAVLEOHAELIKLGEHEKEVYAGLEEKY---KVEATIKLEEKHDVYTKLGEH 1026
555 H---ADPIDEKPV-----GIGHSHN-----YELFKPEE 581
1027 KEELAKLDEGHEVNEYEKNAALNLEENHKEMIKLEEHESASDLVEXLYODE 1086
582 GYAKREGKRYTGEELTVNNILKSTNNONFTLANGOKRVSFPPELEK-KIGIML 640
1087 EV-KNSNKKI---EELTVNIDLNDISICYKQILEEYERNEYV--BEINKLKVONEM 1140
641 VKLITPDGKLEKVSF--VYFGGVYNINAFELDOPYLPQFFKTYIASKDPEVSYDGF 699
1141 KDM--NOKKILEKEIKKLNKLSNIVETKE---NFK----- 1176
700 TVPTSLAYKMASQITFYFPHAGDYTLRVNPOFAVPKGTDALV--RVDFEFGNATLENNY 757
1177 -----NSEMVYNNENKRIIVOSKENISSEVDS-- 1206
758 KVGELK--LPYKLNQGTTRAGKIPYTPMAANAYLNDOSITYVPLEKENDTKRSI 815
1207 KGNILKMTLSLKKERNIFSINDKNESSELYDTI--KSAYINKIEYKKEIEDNGKNI 1263
816 -----LPOFRKNAQOENSL-----DEVEEPTKSEKE----- 844
1264 EDLKNKIILDSNELLNENKMYLTDENNNLKELEIYDNKLNKEKNENNEILANDDI 1323
845 ---KELISGTGNSNSTLEEVPTVDVQEVAAFAESYGNKLENTL--FN-----M 891
1324 IKLKKEISEMDEEKLTKENI---KLKNDIEQINKKEYKKEENLMKFMENINEVYSL 1379
892 DGTIELYIPSEVYKKNNAADTGAPOGNGKNSKSEKSTGVENOPTKRNADSLPE 951
1380 KNOIELEKMKLEENKNTLLAEKRETNMSISNDKNIVNNITL--TSK----- 1430
952 APNEKPVPEPSTNGMLNPGNVSDEPLDPALEAPAVPOVEKLEKTFASYGIGL-- 1009
1431 ---QNNLNR---NVDDKTGDDINCEKNNDQAKISYILDEIKKISMLTGEELNR 1478
1010 ---DSVIFNDG---TIELRLPSGE 1028

```

```

Db 1479 KNSYDEKYNLNLKELKRNKKE 1504
RESULT 10
ID 0904X0 PRELIMINARY; PRT; 2647 AA.
AC 0904X0;
DE 01-MAY-2000 (TREMBLREL 13, Created)
DE 01-MAY-2000 (TREMBLREL 13, Last sequence update)
DE 01-MAR-2001 (TREMBLREL 16, Last annotation update)
DE PUTATIVE ERYTHROCYTE BINDING PROTEIN EBL-1 (FRAGMENT)
GN EBL-1
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2007864; PubMed=10613703;
RA Peterson D.S., Wellem's T.E.;
RT "EBL-1, a putative erythrocyte binding protein of Plasmodium falciparum, maps within a favored linkage group in two genetic crosses."
RL Mol. Biochem. Parasitol. 105:105-113(2000).
DR EMBL; AF131999; A033018.1;
FT NON_TER 2647
FT 2647
SQ SEQUENCE 2647 AA; 304551 MW; AE98F88FD754E300 CRC64;

Query Match 3.98; Score 211; DB 5; Length 2647;
Best Local Similarity 19.38; Pred. No. 0.0091;
Matches 245; Conservative 177; Mismatches 441; Indels 406; Gaps 63;

QY 1 MKFSKTYIAASAVYLSLCAVA-----LNHRSEKNDNRVSVQSQ 46
Db 604 MLKEKYNAGDKICNMLISYADIGDIVRGLDVMDINNRKLEK--FOKITMGGSN 661
QY 47 SSKSENLTPPOVSGKEGIAEQIYIKITDGYVYSHGDHYHYGKVP-----YDA 98
Db 662 RKQNDNRKMKWMEKQNLWSSVMKHIIPKCKTKRHNNNE---KIPQLRLKMGND 717
QY 99 LPSELLKNDPNYQKADIVNEVKGYYIIVD--GKYVYLDAHADVNRTRDEINR 155
Db 718 EFCEEM-----GTEVKOLEKICENKNCSEKCKKACSSYEKIKERKNEYMLQSK-RFDS 771
QY 156 OKOEHYKDN--EYKNSNVAARSGRTTN-----DGYF-----NNA----- 191
Db 772 DKLNKKNLNNKREDSKAYLRSSKQCSNIEFNDETFTEPNKYKKAQMCVENCSSKAL 831
QY 192 ---DI---IEDTGNAYT-----VPH---GGHY-----HYIPKSDLSASE 221
Db 832 KPIKTDVPIETKSKSELSTLDSKNTPNSSGGGNYDROISKRDVNHGPREVYSGE 891
QY 222 LAAAKLAHLAKGNMOPQSOLYSTASDNNTQSVAKGYSKPNK-----SENLO 269
Db 892 KEVPKIDAAVK---TENEFYSNRNDIEGKESKGDHSSPVHSKIDKNEEPORAYSELP 947
QY 270 SLKELVDSPS-----AQRYSDDGYVPDPAKIIISTPAGVAIIPGHDHFIPIYKLS 322
Db 948 KIEKMESSOSIPITHLEAEGQSSNDNDPAVSGESKDVNLHSE-----RIK 999
QY 323 ALBE-----KIAKAVPISGTGTVSTNAKNPNEVYVSSLSGSSNPSLSTSEKL 370
Db 1000 ENEGSVAKITDSSKISIELSKIPSDQNNHSDLSQNA--NEDSNOQNKETINPS--TEKNL 1055
QY 371 ---SSADGY-----IFNPDIYEE-----TATAYVRGDFHPIPSNOIGQ 411
Db 1056 KEIHYKTSDDHSGKISKELEPEKTEESPLTKTESAAI---GKNHESYVSADIFQ 1112
QY 412 PTLN-----NSLATPS-----PSLPINPQ-TSHEK--HEEDG 441
Db 1113 SEIHSNDRIRVSESAYVQSSGSSMSTESTIRDNKDKFTSEDIAPISNGHEKIGSSADD 1172

```

QY 442 YGPDANRIIAEDSGFVMSHGDH-----HYFCKDLTEBOIKAAOKHLEEVY 490
 DB 1173 KSEDSKSIDKSEFENKSSHSDIKOSDNGSTDY---ESLTERSPRG---DLESVSP 1226
 QY 491 SHNGLD-----SLSSHEDYPGN--AKEMDLDKITEEKIA 524
 DB 1227 SSIDMDLKNKSSPPTSSEPHVDSPINSELASASQANADSVQGEKPSKNILRTGDVSEK-- 1284
 QY 525 GIMKOYGVRESIVYKNEKNALITYPHGD-----HHNADPIDEKPYGIGHSHSN 573
 DB 1285 -----EKISVSPWAVTYDEGDKRGQISDSSIHHEIDP-----EKNH 1324
 QY 574 YELFKPREBV-----AKKEGNVYTGEE-----LT 598
 DB 1325 YSFSISEGLEBGEIEKEKEDGSLIPISPERKINDGFRREITVDPSEYVDNDSIDS 1384
 QY 599 NVVNLKNTFNNQNTLANGORVSPFPELEKLGIM-----LVKLTIPD 647
 DB 1385 NVYSEGESEST-----ISSRNGTEGINSELSEKSEHTSVANNRDEHNQENLVSSITQ 1440
 QY 648 GAVLEKVSQVGEVGNIANFELDQPLDQGTFFKYT-----TASRDYPEVSY 695
 DB 1441 ESEPEREKREK---ENADSHSELSISSEVGETIRNDAEASENDKGEDILQSEGOIYT 1497
 QY 696 DGTFFY-----PSTLAKYASQITIFPFAHG--TILKRVN--QFAYPKGIDALV 741
 DB 1498 ETKLEPTVNILOPSTPGTEIHKESSEIDNDNGEVTHTDAPFYRSESEVEVIT 1557
 QY 742 RYVDEEH-----GNALENNRYKGEIKLIPKLNQGTTRAGNKIPVTMANAYLDNOST 796
 DB 1558 NRPDEHEMTKTPSOIIEK--ELGEI-----DSTNQDN-----DEQDN 1594
 QY 797 YVEVPILEKENOTDKPSILPQFRNKAQENSKLDEKVEEPTSEKVEKEKSENGSTS 856
 DB 1595 ST--IFLKNKNEEDGVS--PSTRVMSGSFVSR--NEQIIEEKDQKHIDDTJINPS 1648
 QY 857 NSTLE---EVPYVDQVQEVAKFAESYGKLENVLEPMQGTIELIYSPSEVYIKKNADPT 913
 DB 1649 ENGLKGYGMPN--DST--TSVTTTESPLDVEOMTEPIDG-----KCN--EKN--NIT 1694
 QY 914 GAAPONGENKSENGKSTGTVENOPTENKP---ADSLPEAPNEKVPENSTONGMLN 970
 DB 1695 GEPQESTTIRKOMQPSIN---VNIPELHVAESKIEBAKE---KSMODADKGTIT 1747
 QY 971 PGMVGGSDP 979
 DB 1748 EDITVYEDP 1756
 RESULT 11
 ID 09T274 PRELIMINARY; PRT: 1704 AA.
 AC 09T274
 DT 01-MAY-2000 (TREMELREL 13, Created)
 DT 01-MAY-2000 (TREMELREL 13, Last sequence update)
 DT 01-JUN-2000 (TREMELREL 14, Last annotation update)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR.
 GN MSPI.
 OS Plasmodium falciaparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_Latid-5833;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-HN2;
 RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF062348; AAC72885.1;
 DR InterPro; IPR000561;
 DR Pfam; PF00008; EGF; 1.
 KW Merzoite.
 SQ SEQUENCE 1704 AA; 193762 MW; 385526DDA56FDID CRC64;

Query Match 3.9%; Score 210; DB 5; Length 1704;
 Best Local Similarity 20.2%; Pred. No. 0.0052;
 Matches 248; Conservative 177; Mismatches 411; Indels 394; Gaps 66;
 QY 1 MFSSKRYIAAGSAVYVSLCAVALNHRQEN---KDNRFVYDGSQSSQKSENLT 56
 DB 504 MAFNNNF---DKDYVDKIFSAITYNVEKORVNNRFFSSNNSVYV---QKIKKALSTYLE 557
 QY 57 DOVSQREGIAQIYIKITDQGYTSHGDHYHYNGKYVDA--LFSEEL----- 104
 DB 558 D-YSLRKG-----SEKDFNHYTLKGLGLEDIKKLEIKSENKILE 600
 QY 105 -----LMQPNQLMDADIVN-EVAGGYII-KYDG--KYYVILKDAAHADNRT----- 149
 DB 601 KMFKGLTHSANSLSLESDYKIQVQAVILLIKKIEDLKLELKNRQKLDSTHYVRITYP 660
 QY 150 -----KDEINROKO-----EHVKNERNVNSVA---VARSGRYTNDGYVF 188
 DB 661 QMKPEPYIYLAKKEVDKLEETPRYKMDLKKRQAVLSSITQPIVAASE---TTEG--- 714
 QY 189 NPADITIEDGNAYIVPHGHHYIPKS-----DISASELAARAHLAGKMAQPSQLSY 241
 DB 715 -----GSHETLSQSGETEYEETEETEYTGHTTYITILPPTOPSP 757
 QY 242 SSTA-----SDNNQSAKGSSTKPAK-----SENLOSILKELY 276
 DB 758 PREKAVENSIEHKSNDQALTKYLLKDEFLTKSTYCHRYILVNSSSMDOKILEY 817
 QY 277 D--SPSAQRSESDGLVDFDAKILSRTPNGVALPHG-----DHYHPIPSK-----L 321
 DB 818 NITPEKNKLS---CDLDLDFNIONNIPAMYSLYDSMNNDIQHLPFLYOKEMITYL 873
 QY 322 SALEEK-----TANVPISGTGYST--NAKPNEYSSLSLSPSLTTSKELSSMS 374
 DB 874 HLLKEENHKKILEEQKQITGTSSTSPGNTVYNAQSAKTHSNQOQNAS---TNQ 930
 QY 375 DGYTFPKQIVETATAYVRHGDHNYIPKSNQI--GQPLTN--NSLATPSPSPINPG 431
 DB 931 NG-----VAVSGPAVVEESHDPILTVLSTISNDLGIYSLNLGKTYVNP--LTIST- 981
 QY 432 TSEKHEEDGYFDANRIIAEDSGFVMSHGDHNYFFKD-----LTDEOIKAAOKH 485
 DB 982 TEMEKTYEN-----ILKNDTYF---NDDIKQFVNSKAVITIGLETQKNALNDEI 1029
 QY 486 EEWKTS-----HN---GLDSSLSHEDYPGNAKEMD---LDKITEKTAGIKYOG 531
 DB 1030 KTKLPDLQSLFDLYNRYKTLDRLEFKKKELEGQDKQIKKTLTKQLSKLNSLNNPN 1089
 QY 532 VKRE--SIVVNRKKNAL-----YRPGDHHADPIDEHKPVGT- 567
 DB 1090 VLDNFSVFENKKEKAEIAETENTLENTKILKHYKGLVRYNGE---SSPLTLTSEVSIQ 1146
 QY 568 -GHSNRYELFKPEBGVAKEGNKRYTGE-----LTNVNLKNTSTFNNQNTLAN 618
 DB 1147 TEDNRYANLEKFRYLSKIDKLNNDLHGKKLSFLSSGHHILTEK--EVINKNNT- 1202
 QY 619 GOKRVSEFPPELEKLGIMLVKL--ITPDGVL-----EKVSGVYFQBGVGNIAN 668
 DB 1203 -----GNPSSENNK--VNEALKSYENFLPAKVTVTVPPODPTPPLSLVSRVSGSG 1254
 QY 669 FELDQPYLPG-----TFYTTASRDYPEVSYDGTFTVPTSLAYKASQITIFPFAAGT 723
 DB 1255 STKEQIPTSGLLTLEQVAVOSQNTDEED--DSLVLV-----ITGESEDNDE 1302
 QY 724 YL--RVNPOPAVYKGTDALVRVDFEFGNAVLENRYKGEIKLIPKLNQGTTRAGNKIP 782
 DB 1303 YLDQVYTGALISTYMDNITSGF-----ENEDVYIYK--PL--AGVYRSLKKQJE 1348
 QY 783 ---VTFMANA-----YLD-----NOSTYIYVPILEKENOTDPS 814
 DB 1349 KHLTFNMLNLNDILNSRLKRRKFLDVLSEDLQMFHISNRYIE--DSFKLLNSEQKVT 1407
 QY 815 ILPQFRNK-----AOENSKLDEKY-----EPPKISEKVEKEKSENGSTNSNLT 860


```

Db 1408 LKSYTYIESYENDIKFQEGISYEKYLAKYKDDLESIKYIKR--EKEPSSPPT 1464
Qy 861 EEPYVDPOEVAKFA-----ESYGMKLENVLFNM-----891
Db 1465 PRSPATDQKESKPLPLTNIEYLNVLNKKIDYLNLAACINDCNDVEDEAVKTT 1524
Qy 892 -----DGTIELYLPSEGVAKKNNADFTGEAPQNGENKPSNGK-VSTGTVENOPT- 941
Db 1525 KLSDDKAIDKIDLF-----KNTNDFEALIKKLINDTKDKMLKSTGLGVNKPRT 1576
Qy 942 -----ENKPADSLPEAPNE--KPYKPPNS 963
Db 1577 IISKLEKGFQDMLNISOHCYKQCPNS 1606

RESULT 12
ID 025922 PRELIMINARY; PRT: 1720 AA.
AC 025922;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PRECURSOR OF THE MAJOR MEROZOITE SURFACE ANTIGENS.
Plasmodium falciparum (Isolate NF54).
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI TaxID=5843;
[1]
RP SEQUENCE OF 1069-1720 FROM N.A.
RC STRAIN-NF54;
RA Tolle R., Bujard H., Cooper J.A.;
RL Exp. Parasitol. 0:0-0(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-NF54;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP STRAIN-NF54;
RC MEDLINE=96123395; PubMed=8577332;
RA Pan W., Tolle R., Bujard H.;
RL "A direct and rapid sequencing strategy for the Plasmodium falciparum
antigen gene gp190/MSA1."
Mol. Biochem. Parasitol. 73:241-244(1995).
RL EMBL: 235327; CA84556.1;
CR InterPro: IPR000561;
CR Pfam: PF00008; EGF; 1.
CR Signal: Metazoa;
CR SIGNAL 1 19 POTENTIAL.
CR SEQUENCE 1720 AA; 195725 MW; 717B2FD1E637A8A3 CRC64;

Query Match 3.8%; Score 205.5; DB 5; Length 1720;
at Local Similarity 20.1%; Pred. No. 0.0094;
atches 248; Conservative 171; Mismatches 413; Indels 403; Gaps 65;

```

```

Qy 189 NPADIEDTGNAYIVPHGHYIPKS-----DLASSELAAAHLAGNMPSQIST 241
Db 730 -----GSHFTLLSQSETEVTEETEEVGHVITVITLPLPTQSP 772
Qy 242 SSTA-----SDNTQSVYAKGSKSPAK-----SENQSLKRLX 276
Db 773 PREVRVENSIEHKNSDQALTKVYLLKIDELTKSYICAKYILVNSNDKQLLEVY 832
Qy 277 D-SPSQRYSSEGLVFPKATISRTPNGVALPHG-----DHYHIFPSK-----L 321
Db 833 NITPEEEMKLS-----CPDLLEFNQNNIPAMYSIDXSMMNDLHLEFELYOKMITYL 888
Qy 322 SALEEK-----IANNVPLSGTSTVST--NAPNEVSSLSGLSNPSSLTTSKLSAS 374
Db 889 HKLKEENHIKLLIEQOKOITGSTSTSPGNTVNTAOGATHSNSQOOSMNS--TMTQ 945
Qy 375 DGYINPNDIYEETATATVIRGDFHFIKPSNOI--GQPTLPN--NSLATSPSLPIRG 431
Db 946 NG-----VAVSGPAAVESHDPLTVLISINDLKGIVSLINLGNKTRVPRP-LTIST- 996
Qy 432 TSHKHEEDGYGFDMRITIAEDSGFVASHGDNHYFPKID-----LTEDQIKAAOKHL 485
Db 997 TEMEFYEN-----ILKNNDYF--NDIKQFYKSNKVTGLTFQKNALNDEI 1044
Qy 486 EEWKTS-----HN-----GDSLSHEDYPGNAKEMKD--LDKTEKXIAGIMKOYG 531
Db 1045 KKLKDTQLSFDLYKRYKILDRLENKKELQDQKQIKLTLLEQESKLINSINPDN 1104
Qy 532 VKRE-SIVNKEKNLI-----YPHGDHNDPIDEHKPGCI- 567
Db 1105 VLQNSVFNKKKEALIEETENTLTKILKHYGLVKNYNGE--SSPLTJSEVSIQ 1161
Qy 568 -GHSNTELEKPEEGVAKKGNKYTGEE-----LTVVNLKSTNNONFTLAN 618
Db 1162 TEDVNALEKRYVLSKIDKLNQNLHLQKKLSPLSSGHHITLTK--EVAKNNYT-- 1217
Qy 619 GOKRVSFSPPELEKILQINMLVKL--ITPDGKYLEVSGRVGEGVNIANFELDQY 675
Db 1218 -----GNSPENNKK--VNEALKSENFLEAKVTYVTP-----PGPD 1254
Qy 676 LPQGTFTT-----ASNDYPEVSYDGFYPTSLAYMAAQITFPPHADDTYLRVAPQ 730
Db 1255 VTPSPLSVAVSGSGSTKEETQIPPSG-----SLTPELQOVVOLQNTDEDDSLVLP 1308
Qy 731 FAVRQGD-----ALVRVDEFGNAVLENNYKVEIKLIPRLNGTTRTAGNK 780
Db 1309 FGESENDDEYDQVYTGALISYTMDNISG--FENEYDVIYIK-PL--AGVTRSLAKQ 1361
Qy 781 IP-----VTFMANA-----YLD-----NOSTYIVAEVPLLEKENQTDK 812
Db 1362 TEKNIFPNLNLNDILNSRLKRRKYPDLVLESDDLMOFHHISNETIIE-DSFKLLNSQK 1420
Qy 813 PSILPQFRNK-----AOENSKLDEK-----EKPQSEYKVEKKLSETGNST 855
Db 1421 NTLKSYTYIESYENDIKFQEGISYEKYLAKYKDDLESIKYIKVEKKKPPSSPPT 1480
Qy 856 SNSLLEVPYVDPOEVAKFA-----ESYGMKLENVLFNM-----891
Db 1481 PPS-----PAKTQKESKPLPLTNIEYLNVLNKKIDYLNLAACINDCNDVEDEA 1535
Qy 892 -----DGTIELYLPSEGVAKKNNADFTGEAPQNGENKPSNGK-VSTGTVE 937
Db 1536 HVKTKLSDKAIDKIDLF-----KNYDEALIKKLINDTKDKMLKSTGLGVQ 1587
Qy 938 NOP-----ENKPADSLPEAPNE--KPYKPPNS 963
Db 1588 NFPTIISKLEKGFQDMLNISOHCYKQCPNS 1622

RESULT 13
Q9LBG3 PRELIMINARY; PRT: 1653 AA.
AC Q9LBG3;

```

01-OCT-2000 (Tremblrel. 15, Created)
 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE PAA.
 GN Streptococcus criceti.
 OS Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1333;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B49;
 RA Tamura H.; Kato H.;
 "Cell surface antigen I/II - Streptococcus cricetus."
 Submitted (Apr-2000) to the EMBL/GenBank/DBD databases.
 EMBL; AB042339; BAA95000.1;
 DR InterPro; IPR001899;
 DR InterPro; IPR002965;
 DR Pfam; PF00746; Gram_pos_anchor.1.
 DR PRINTS; PRO1217; PRICHXTESEN.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
 SO SEQUENCE 1653 AA; 18062 MW; 93F28E215BF598DE CRC64;

Query Match 3.7%; Score 198.5; DB 2; Length 1653;
 Best Local Similarity 18.6%; Pred. No. 0.021;
 Matches 242; Conservative 166; Mismatches 435; Indels 459; Gaps 63;

2 KFKKTYIA--GSAVYISLSICAVYALNQHRSQENKDNRYSYVD-----GSQSSQKSENT. 54
 13 KIKSLGALGTAIYVSA-----GQRLADMTSTSAVDTAIVGTETGNPATNL. 66
 55 ---TPQVSKQEGIOAEQIYKITTDOGYTSHGDHYHYNGKPYDALSE----- 102
 67 PERQADSSSQAEKSOAOA-----EOKTGMPPVAVATTELDKAVKSA 108
 103 -----ELKKDPRYQKADIVN-EYKGGY-----IIRKDGYYVYLKDAADNV 147
 109 EAGVTVSODTIVKGVGVSQEDKESGETKADYSKQAEITKTTETV-----KADVA 161
 148 RTQDEINRQKQEHVKNKENVAVASQGRYTTNDGYFNPADITIEDGNAYIYVHG 207
 162 ANQSENRINQENAAKQAYEODLANKAKEVEITNK-----NQ----- 201
 208 HYHYIPKSDISA-----SELAAR-----AHLAKNMOPQSUSYSTASDNNQSY 253
 202 -----AKADYEAKIAGYOKELAAVQOANSNDQAAYAAKADAYEKELARVQKA---NADAK 253
 254 AKGSTSKPANKSENQ-SLILKELYDPSAQRYSQGLVDPKAITISRTNGVAIFHGDH 312
 254 AAYEQAVANTATNEQIKAMAAIQORNAQAKADYEAKIAGYOKDLAAVQAGNAHEADY 313
 313 YHPIPSKLSALEEKIARWVPISGTST-----VSTNAKNEVYSSLSNPSSTLTS 367
 314 -----QAKTAYQDELARVQOANAAKTAAYEQAVANTATNEQIKAMTAIQORNAQAKA 368
 368 -----KELSSASDGYIFNPDI-VEETA-----TAYTVRRGDHP 400
 369 DYBAKLAQYONDLAIVQAGNAHEADYQAKKTAYEDELARVQOANAAKAAKAY----- 420
 401 HYTPKSNQIQOPLTPNNISLATPSPSLPIFGTSHKEHEDGYTFDANRIITAEDESGFWS 460
 421 -----EQAVANTATNEQIKAMAAIQORNAQAKADYEAKIAGYOKDLAAVQAGNAHEADY 473
 461 HGDHNYEFKDKLTERQIKAKQKHEVEKTSHNGL-----DSLSSHEO----- 503
 474 EAD---YQAKKTAYDEELARVQKANDAKATAYQAVKDNQAKNAETIAEAAAIKORAAA 530
 504 ---DYPSN-AEMKDL-----DKRT--EEKIAGIMKQYGVNR-----ESTIV 539
 531 KADYEAKIAGYOKDLAAVQAGNAHEADYQAKKTAYETELARVQKANAADAKAAYDQAVKD 590
 540 NKEKNALITYHG----- 551

DB 591 NQAKNEIAENNAIQRNAAKADYEAKIAGYOKDLQYOKDFADYQSKRAYEDQA 650
 QY 552 -----DHHHAD-----PIDEKPPVIGHSNLYLFPREEGVAK--EGNKVYTGEL 597
 DB 651 YKAALAELEKHKNDGYLTPSPESQPF-----IFQENKATLETISGDKVYNADEF 700
 QY 598 ---TNVYNLLKNSF-FNNQFTLA-----NQKRVSPSPPELEKKGIMLVYLITPDK 649
 DB 701 YAEADKLDYIKADLTLENTERRRAYEHPFGGIDSDADWPK-----RVFLPKRK 749
 QY 650 -VLEKSGRVFEGEGVGNINAFELDQPYLPQGFYKYLASRD-----YPEVSYDGFYV 701
 DB 750 PYVARIS-----NLKNSLNGKKIASAEITYTL--KDTSLNADKIPAMLYD----- 794
 QY 702 PISLAYKMASQITFYFPAGDYTLRNPQFAYVKG-----TDALVYVDFEFGHNALENN 756
 DB 795 PL-----QTIYFEFFGSGSIDVDIKFYDEDKNKIDATGALDSESLNRHGMPQYN 846
 QY 757 YKGEIKLPIPKINQSTTRAGNKIPVTFMANAYLDNOSTYIYEVPILEKNO----- 809
 DB 847 -GINTLE---KVNHYT-----GTALFIN--GSSVKDNGDGLVATNTYRKANSRESNE 895
 QY 810 ---TDKP-----SIL-----POFKRNKAQENSK-----IDKYEKPTSEKYEKELSET 851
 DB 896 WDTNTPVMWYGAIVGKQGPDIKFTLSAENAGIWFARDSKI-----KAKOI 943
 QY 852 GNSTNSLTLEVPYTPDPOEKVAKFAESTGAKIENLVFNDGTIELYLSGGEYIKKNMAD 911
 DB 944 PKNPMS--EPFPVAPQ-----PLE-----PYEVEK-----D 971
 QY 912 FTGEAPQNGENKSENGKYSTGVENQPT-----NKPADSLPEAPN--EKPVPENST 964
 DB 972 LEFVPVPEPNTDPEPTPTKPPDQPEPNKPDPEPNYEKESSEAPPIEPNTDPEPTPTK 1031
 QY 965 DNGMLN-----PEGNVGSDPMLDPALEAAVDPVQDEK----- 998
 DB 1032 DQEPNKPDEPNYEQGPAPAVEPKYD--EPTPPVDYVEKIKIDPVVPTVRYHYK 1089
 QY 999 -----EKTASYGLDQSVIFNNDGTIELR-----LPBG 1027
 DB 1090 LAIQPVTAKIKNSDMLDIDKTLVAKOSTVFKQLNADLPAG 1131

RESULT 14
 09LC00 PRELIMINARY; PRT; 1698 AA.
 AC 09LC00;
 ID 09LC00;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE HYPOTHETICAL 181.9 KDA PROTEIN.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCTC10442;
 RA Ito T.; Hiramatsu K.;
 RT "Comparison of three types of S.rmeec."
 RT Submitted (Oct-1999) to the EMBL/GenBank/DBD databases.
 DR EMBL; AB033763; BAA86640.2;
 KW Hypothetical protein.
 SO SEQUENCE 1698 AA; E209D67161F9D84B CRC64;

Query Match 3.7%; Score 198; DB 2; Length 1698;
 Best Local Similarity 18.9%; Pred. No. 0.024;
 Matches 228; Conservative 144; Mismatches 465; Indels 372; Gaps 50;

2 KFKKTYIAGSA-VIYISLSIC-----AYALNQHRSQENKDNRYSYVDS-----OS 47

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:27:16 : Search time 46.65 seconds
(without alignments)
1696.578 Million cell updates/sec

File: US-09-471-255-2

fect score: 5406
Jence: 1 MKRSKRTIAGSAVIVSLSL.....IELRLPGEVYIKKMLSDPIA 1039

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	921.5	17.0	822	2	T46758		hypothetical 92.4K
2	230	4.3	2485	1	H71621		serine/threonine-s
3	226.5	4.2	1701	2	A54498		major merizolite su
4	218	4.0	1873	2	T30944		surface protease pr
5	215.5	4.0	1701	2	A26868		major merizolite su
6	214	4.0	1939	2	T18372		repetit organellar
7	205.5	3.8	1185	2	A42404		collagen adhesin -
8	205.5	3.8	1726	1	SAOQM		major merizolite su
9	203.5	3.8	1849	2	C41859		Iga-specific metal
10	202.5	3.7	1726	2	A45948		major merizolite su
11	196.5	3.6	1271	2	A45555		glutamate rich pro
12	196	3.6	5005	2	F82884		hypothetical prote
13	193	3.6	3724	2	T18427		hypothetical prote
14	190.5	3.5	4549	2	T20771		hypothetical prote
15	190.5	3.5	4667	2	T20774		hypothetical prote
16	187	3.5	1080	2	T43164		lacta protein - Ll
17	187	3.5	4688	2	F82885		hypothetical prote
18	186.5	3.4	1570	2	T18272		1-phosphatidylinos
19	186	3.4	1231	2	T28681		rib protein - Stre
20	185.5	3.4	1071	2	E85343		hypothetical prote
21	185	3.4	2195	2	S61103		SEC16 protein - ye
22	184.5	3.4	1466	2	A36426		SP22 protein - ye
23	184	3.4	3381	2	T43389		verecan precursor
24	182.5	3.4	1023	2	T48997		epsin-like protein
25	181	3.3	1879	2	UW0059		mcpd protein - mo
26	180.5	3.3	1883	2	G86643		hypothetical prote
27	180.5	3.3	5170	2	T15348		hypothetical prote
28	179.5	3.3	1315	2	T28679		fibrinogen-binding
29	178	3.3	1436	2	S57238		forced protein 5.4

30	178	3.3	1449	2	S57237		forced protein 5.6
31	177	3.3	1702	2	A41859		Iga-specific metal
32	176.5	3.3	1139	1	E64234		cytadherence-acces
33	176.5	3.3	1532	2	A26036		Iga-specific metal
34	176.5	3.3	1566	2	A43607		cell surface antiy
35	176.5	3.3	1658	2	S55101		hypothetical prote
36	176	3.3	3147	2	T18674		hypothetical prote
37	175.5	3.2	2401	2	T28676		rhoptery protein -
38	175.5	3.2	2748	2	S57976		nuclear migration
39	174.5	3.2	1577	2	A35140		hemolysin A precu
40	174.5	3.2	1807	2	S03124		vitellinogen A2 pr
41	174.5	3.2	2294	2	I67630		protein tyrosine p
42	174.5	3.2	2466	2	I67629		protein tyrosine p
43	174.5	3.2	2490	1	A54971		protein-tyrosine-p
44	174	3.2	1631	1	SAOXM		major merizolite su
45	173.5	3.2	2722	2	T20532		hypothetical prote

ALIGNMENTS

RESULT 1
T46758
hypothetical 92.4K protein - Streptococcus agalactiae
C1Species: Streptococcus agalactiae
C1Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C1Accession: T46758
R1Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heymann, J.; Schultze, N.; L
Infected. Immun. 67, 871-879, 1999
A1Title: Lmb, a protein with similarities to the Lrat adhesin family, mediates attach
A1Reference number: 224091, PMID:99115568
A1Accession: T46758
A1Status: preliminary; translated from GB/EMBL/DBJ
A1Molecule type: DNA
A1Residues: 1-822 <SP>
A1Cross-references: EMBL:AF062533; NID:94249622; PIDN:AAD13797.1; PID:94249624
A1Experimental source: strain R268
C1Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match	17.0%	Score 921.5	DB 2	Length 822
Best Local Similarity	26.9%	Pred. No. 5e-38		
Matches 267	Conservative 152	Mismatches 305	Indels 267	Gaps 31
QY	1	MKRSKRTIAGSAVIVSLSCAALNQHSGENKDNRRSVYVGSQSSQKS--ENLTPDQ	58	
DB	1	MKRTYIGISVAAILATHIGSYOLKHHGGLATKQNLAVIDSQKVAAPRTKNTMDQ	60	
QY	59	VSOKEGLOAEQIVIKITDGGYVTSRGDHYHYNGKVPYDALFSEELMDPNYQLDADI	118	
DB	61	ISEEDISAEQIVKITTDDGYVTSRGDHYHYNGKVPYDALFSEELMDPNYHFGOSD	120	
QY	119	VNEVKGYYIKYDKYVYVYLKDAHADNVTKDEINRQKQEHYD--NEKYSNVA	172	
DB	121	INRILGGYIKVNGYVYVYKPSKRNKIKTKQIAEQVAKGTEKKEKGLAQAVALHSE	180	
QY	173	-----VASQGRITNDGVFNPADIIETGNAYIVPHGCHHYIPKSLSELAAR	226	
DB	181	EVAANVEARKQGRYTTDDGTFESPTDIDLDGAVYVPHNHHYIPKDLSPSELAARQ	240	
QY	227	AHLA---GNKMPQSLSYSTASDNNTQSVAKGSTKSPAN-----	263	
DB	241	AYTSQKQSGARPS--DYRTPAPGRKAPIPRYTNPQGGHQPNDGYPAPRPNDAS	298	
QY	264	-----KSENLQSLLELYDPSAQRYSSESDGLVFPDAKTIISRTPNCAVPHGDHY	314	
DB	299	ONKHQDERFKGTFKFKEDQLHRLDLKYRHYVEDGILFEPTQVYKSNAGYVVPDHDHY	358	
QY	315	FITYSLALAEKIAKAVNPISGSGVSTAKRNEYVSSISGSSNPSSLTSSKELS---	371	
DB	359	IIPRSLSPLEMLADRYLAGOTDDNDS-----GSDHSKPS-----DKEYTHRF	402	
QY	372	-----SASDGYIFNPKDIVEETATATAYIVRHGDHPHYIPKSNQIQOPTL	414	

```

2b 403 LGHRTKAVGKGLDGPVPTSDAAYVSKESISVDSKGTAKHGDFHFI -GEGLEQVEL 461
2y 415 PR-NSLATPPSPSLPINGTSHKEHEEDGCPD---ANKRIAEESQFVMSHGDNHYF 468
2b 462 DEVAWVAKAGQDELVALDQEOGKERPL-FDTKRVSRKTKDKQKGYIMPKQKIDFY 520
2y 469 FKRDLETOIKAAQKHELEEVTSYSHGLDLSLSHEQDYGNKEMKDLDKTEETAGIMK 526
2b 521 ARYQDLDQIAFAEBELMLKDKKHRYDI-----VDTGLEPRLANDVS 563
2y 529 QYGVKRESIVNKKENALITYPHGDHHPD---IDENKPVGISHSNYTELKEE--EGV 583
2b 564 SLPMHAGNATYDTGSSFYI-PRIDHIVVFPYSWILRNQIATL-----KYVQHPVPRDV 617
2y 564 AKKEENKTYTEBELNYYNMLKNSFNNONFTLANGKQARVSFPPELEKTLGIMLYKL 643
2b 618 WSKPQHE-ESGSGVIVNVPDLDRKGMFMWQI-----IHSAEVQKALAE--RF 663
2y 644 ITPDGKVLKESGKVGFGVGNIAFELDQYPLPGQTEKTYIASNDYEPVYDGFYVPT 703
2b 664 AAPDGYTD-----PROVLAKETP-----YKDGSEFSIR 693
2y 704 SLAYMAASOTIYPPHAGDTYLKVPQAVKRGIDALYVDERHGNAYLENNYVGSIK 763
2b 694 A----- 694
2y 764 LPRLNGTETRTANKIPYTFMANAYLDNSTYVEVILEKEN-----QTDKSLIPQ 818
2b 695 -----DGSISRTINKDL-----SQAEWQAOELLAKKRNAGATITDKEPEEQQ 738
2y 819 EKRNKAQENSKIDEKVEEPEKTSKYEKKEKSETGNSNSTLEEVPTD---PVQEKVA 874
2b 739 --ADKSNEN-----QQPSEASKKEKE-----SDDFISLPGVGLRATLEHDIN 780
2y 875 KFAESYGMKLENVLEFNDQITELIYPSGEVI 905
2b 781 QLAQKANIDPXYLIFQPEG-VQFYKNGELV 810

RESULT 2
471621
serine/threonine-specific protein kinase (EC 2.7.1.-) PFB0150c - malaria parasite (Plasmodium falciparum)
Species: Plasmodium falciparum
Date: 10-Sep-1999 #sequence, revision 10-Sep-1999 #text, change 10-Sep-1999
Accession: H71621
iGardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Arevalo, L.; Koonin, E.V.;
science 282, 1126-1132, 1998
Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
Reference number: A71600; M01D:99021743
Accession: H71621
iStatus: preliminary; nucleic acid sequence not shown; translation not shown
iMolecule type: DNA
Residues: 1-2485 <GAR>
osr-References: GB:AE001376; GB:AE001362; NID:93845108; PIDN:AACT71820.1; PID:9384510
iGenetics:
iGene: PFB0150c
iSuperfamily: malaria parasite serine/threonine-specific protein kinase PFB0150c; prote
iKeywords: Phosphotransferase
i2087-2352/Domain: protein kinase homology <KIN>

Query Match 4.38; Score 230; DB 1; Length 2485;
Best Local Similarity 18.08; Pred. No. 0.0042;
Matches 209; Conservative 154; Mismatches 397; Indels 400; Gaps 47;

2y 12 SAVIVSLCAVYALNQHSQENKDNRRVSYDGSQSSQKSENLTPDQVSOKEGIAQDIY 71
2b 553 SISLINNVFFYKKRKSSNNNNNNNNNIS--SSSSSSKKNHV---INK- 598
2y 72 IKIDQGYVYSHGHYHYNKKVYPYDALFSELLM----- 106

```

Dh	599	-----ISSYNIHYERKDSFENLFFKEKILPSKDOTCVFNERQDOLPEKSMEHK	650
Qy	107	-----KDRPYOLKADLYNEVKYGGIILYVDGKRYVYLAKDAHADN	146
Dh	651	CVSFFNTSDDISSHSSVNNKEPFFALKNSIHIFPENNIILYTSQKSPNHQODEKTVL	710
Qy	147	VRTEDEINROKOEHVKNONEKSVNVAARASQGYRTINDGVFNPADI-----IEDTG	198
Dh	711	LKKKEIN--DKNTSSCLINHTT-----YTLONGVNNKLNMLGTRDSTYKIDEN	761
Qy	199	NAYIVPGGHYHYPKSDLSASBLAAKAHLAKKNOPQSOLSTSYSDNNTQSAKGSF	258
Dh	762	NMLKECVGNNDNNKKKKKKKLSF-COIIDNTLP-----YESQEKENNNIKSMOI	816
Qy	259	SKPANKENIOLSLKELYDS--PSAQRYSDDLYDPDAKIIERTNGVGAIRHGDHYH	316
Dh	817	FNYYKRSNLTNNLSNRDSTVDMHKNYSEYI-----NIQRINKYELSNKRIRNYK	870
Qy	317	PYS-----KLSALEKIAIARWPI-----	334
Dh	871	LYSDELIFKYSLKAKKTIIDNISNMDEVYTKNMINEKISKMDILYPODKKNSLMSCP	930
Qy	335	-----SGTGSTVSTNAKPN-----VSSLSGLSNPSS	363
Dh	931	VLIENNISREBENKNSVILNKKKNEMFNCVGRLCHMKKNNONDIYDQIKKNEBE	990
Qy	364	LTSKEISASDGYIFNPKOIYEBTAAYIVRGDHPHYPKSNOIGLOPLRPNLSATPS	423
Dh	991	ITKDEYISRBEKKKYSK-----CIRNFDIKY-----EYV-----	1022
Qy	424	PSLPINPOTSHHEKHEDEGYFDANRIIADE-----SGFYMSGDHNNHYFKKD	472
Dh	1023	-----LSYHTLEDCKKKNDMMNLIIDMNEAIIETVGVYLNIIILDRKONS--RKO	1072
Qy	473	LTEQITAAQKHLE--VKSHGSLSHEDDYGNKAKKE--DLDKIIEKIKIAGIMQ	529
Dh	1072	MEKEMEKEKMEKEMEKEKMEKEMEYEME-----KELNENNNNNNNNNENKMN	1120
Qy	530	-----YGVKRSIYVNNKEKAIYIPDCHHADRIDEKH--PVGIGSHSNY	574
Dh	1127	EINIYKNNELIYDNDKLELYNEBK-LIIPRYES--DYHKMMNSININCKDYI	1180
Qy	575	-----ELFKPEEGYAKKE-----GNKYTT--GEELTYNNYLK	605
Dh	1181	NNILKEYVDSCLAQKRENIIFRLPLNMLKDKYWKRFNIKNNIKITIIHNEEMKRIYQIIN	1240
Qy	606	NSTFNNOFT-----LANGOKRVSFSPPELEKCLINMLYKILITPDGVYLEKVSGRYPG	660
Dh	1241	KNWPIYNFRYENFLIN--HLTYNPK-----NOLFKL-----SYKYSM	1278
Qy	661	EGVGN--IANFELDOP-----LPQO--TFKTIASKDYPE--VSYDGTVPYLSLAY	707
Dh	1279	NNIRNLYIAKHHNNNDYANKLYNONIYTLKQVANIIDNHICKKGGGLDIYANNISK	1330
Qy	708	KMASQITFYPHAGDYTLRVPQFAVPKGTALVRYDEFHGNAYL-----ENNY	757
Dh	1339	ECKNR-----KDKETL--NKIFHKKKKDAFFINDEIGSNDYMDIKKYSMDENY	1380
Qy	758	KVGEIKLPIPKLNOGTRTAGNKIPVTPANATLNOQSTIYEVPLIEKNOTDKPSLP	817
Dh	1390	KLNE-----KNNISMSDEMI-----PILNSEHONNPPSOP	1422
Qy	818	QFRKNKQENSKLDEKVEBPKTSEKYEKELSTGTSNSTLLEPVYDPQOEYAKFA	877
Dh	1423	NMLEK-----STYIDLNLIDSNSMDTTEEKYFVNNEN-----DLFNTYKMFN	1468
Qy	878	ESYGMKLENYLF-----NMDGTIELYLPSEVIKKNMADETG-----EADPGNGEN	923
Dh	1469	FSKGMLENNKFFPNVSNEDOVSEFF-----KNNMLFRELKNSNLSIKLESYKNSNN	1520
Qy	924	KPSBNGKVSIGYENOPTER	943

Db 1521 CSNNKGDNDIGNMNNNTN 1540

RESULT 3
major mezozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (154498)

C:Species: Plasmodium falciparum

C:Date: 28-Oct-1999 #sequence_revision 28-Oct-1999 #text_change 09-Jun-2000

C:Accession: A54498

C:Reference: M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.;

Mol. Biochem. Parasitol. 27, 293-302, 1988

Title: Variation in the precursor to the major mezozoite surface antigens of Plasmodium

Accession: A54498; MUID:88142999

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-1701 <P>

A:Cross-references: GB:M9143; NID:9160412; PIDN:AAA29653.1; PID:9160413

C:Superfamily: major mezozoite surface antigen

C:Keywords: surface antigen

Query Match 4.2%; Score 226.5; DB 2; Length 1701;

Best Local Similarity 20.2%; Pred. No. 0.0034;

Matches 250; Conservative 175; Mismatches 406; Indels 405; Gaps 66;

1 MRFSTKIAGSAVIVSLCAVALNQRSOEN---KNNRYIVDGSOSOSKSENLP 56

500 MEFNNF---DKVVDKISARTYVNEKORNNKSSNNSTYV---OKLKALSTYL 553

57 DOVSOKEGIAEQIVIKIDGVVTSBGHGHYNGKVPYDA---LFSEEL----- 104

554 D-YSLKGI-----SEKDFNHTYTKTGLEADIKKLTETIKSEKNIIE 596

105 ---LAKPNTOLKADADIVN-EVKGYYII-KYDG---KYVYVLKDAHADNVT----- 149

597 KMFKGITHSANSLKESVDIVKLYOVKVLIIKKEDEKLELEKNNQKDSIHVPITVP 656

150 ---KDEINROKQ-----EHVKONEKYNVNA---VARSGRYTTNDG--- 185

657 QNKPEPYLLIVAKKEDEKLEFPKVKMDLKEQAVLSSTIOPVLAASE---TTEGGHS 713

186 ---YFNPADIIEDGNATIVPHGGHYHI---PKDSLSASELAAKAHIA 230

714 THTLSGSETETEETEETEET-----VGTTVTITLPKESAPREKAVVNSIE 765

231 GKNMQSOLSYSTASDNMTQSAVAKSTSPARK-----SENQSLKEL 275

766 HK-----SNDNSQALTKIVYLKAKDEFLTKSYICHKYLIVSSSDQKILEY 812

276 YD-SPAQARYSESDGLVFPDAKISRTFPGVALPHG-----DHYHTPYSK----- 320

813 YNLTPEEMELKS---CPDLLEFNIGNNIPMAYSLYDSMNIDLOHLEFELYQKEMITY 868

321 LSLALEK-----IAWVPISTGSGTVST-NAKPNVYSSLSGSLSNPSSLTISKLSA 373

869 LHLKLEENHIKLEBKOKITGSSSTSSPGNTVTVAQSTHNSONQOSMSS---TWT 925

374 SDGIYNPDIYVETATVIVRGDHPHYIPKSNQI-GOPTLPN---NSLATPSPSLIP 430

926 QND-----VAVSSPAYVEESHDLPLVLSISNDLKGIVSLNKNKTVPP-LEIS 977

431 GTSHEKHEBDGYGFANRIIAEDSGFVNSHGHNYFFPKD-----LTEQIKAAQH 484

978 -TEMEKFEYER---ILKNNDYF---NDIKOFVNSKSVITGLTEQKNALND 1024

485 LEEVYKTS-----HN-----GLDLSHEDVYGNAKEMD---LDKTEKTAGIMKOY 530

1025 IKKLKDTQLSPDLKYRYKLKIDRLFNKKKEIGADQKQIKKILLKEQLESKLSNINNP 1084

531 GVKRE-SIVNKEKNII-----YPHGDHHDADPIDEHKPVGI 567

1085 NVLQNSVFPNKKKEFLAETENTLENTKILKHGYGVKTYNGE---SSPLATISEVSI 1141

QY 568 --GHSHSNTELEKPEEGVAKKEGNKRYTGE-----LTVNVLNKSPPNNONTFLA 617

Db 1142 QTEEDYANIEKRALSKIDGKNDMLHGGKTLSPSSGHLITELK-EVYKNNYNT-- 1198

QY 618 NQKRVSPSPPELEKIGINLVTL---TTPDGVYLEKVSQKRVGEGVGNIAFELDOP 674

Db 1199 -----GNSPENNKK--VNEALKSYENFLEAKVTVTVTP-----PQP 1234

QY 675 YLPQGTFTKI-----ASKDYVEVSDGFTFTPTSLAAYMAQOTIFYPHAGDTLRYNP 729

Db 1235 DVTPEPLSYRVSGSSGSTEKEQIPTSG-----SLTLEQOYVOLONTDEEDSLVLP 1288

QY 730 QPAPVKGTD-----ALRVDEFHGNAYLENNRYKGEIKLPIPLNOGTTTAGN 779

Db 1289 IFGESEDNDEYLDQVVTGALISVTMDNLSG---FENEYVITLK-PL---AGVYSLAK 1341

QY 780 KIP---VTEMANA-----YLD-----NOSTIYVEVPLEKENOTD 811

Db 1342 QLEKNITFENLNDILNRLKRRKFFLDVLESIDLMOFKHISNEYIIE-DSFKLINSQ 1400

QY 812 KPSILPQFRNK-----AOENSKIDEV-----EKPTESEKYEKELSETGNS 854

Db 1401 KNTLKSRYKIESEVENDIKFAQEGISTYKELAKYKDLSEIKYIKKEKEKFPSSPT 1460

QY 855 TSNSTLEVPVDPQVEKAKFA-----ESYGMLEVLFFNM----- 891

Db 1461 TTPS-----PAKTEQKSKSLPLPLTNIETLKNLVNKHIDYILNKKAKINDCENXE 1515

QY 892 -----DGTIELYLPSEGVIRKNNADFTGEAPQNGENKPSENGK-VSTGV 936

Db 1516 AHVKITKLSDLAKIDKIDLF-----KTNIDFEAKIKLINDDKKMLGKLLSTGLV 1567

QY 937 ENQPT-----ENKPADSLPEAPNE-KPYKPENS 963

Db 1568 QNFPTIISKLEGRQDMINISQHCVRKQCPENS 1603

RESULT 4
T30944

surface protein precursor - Enterococcus faecalis

C:Species: Enterococcus faecalis

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: T30944

R:Shankar, V.; Baghdadyan, A.S.; Huyke, M.M.; Lindahl, G.; Gilmore, M.S.

Infect. Immun. 67, 193-200, 1999

A>Title: Infection-derived Enterococcus faecalis strains are enriched in esp, a gene

A:Reference number: 220943; MUID:99081742

A:Accession: T30944

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1873 <SH>

A:Cross-references: EMBL:AF034779; NID:93873186; PID:93873187; PIDN:AAD09858.1

Query Match 4.0%; Score 218; DB 2; Length 1873;

Best Local Similarity 20.8%; Pred. No. 0.011;

Matches 215; Conservative 122; Mismatches 377; Indels 322; Gaps 49;

QY 91 NGKVPDALFSEELMKDNYOLKADIVNEVGGYIIKVDGRYYLLDAADAD-----N 146

Db 593 NKRLPDAEYSKTEEDTGNVTDSTGTYVAKNGSTFPVDVEAFKASQAMENDATYVP 652

QY 147 VTKDEINROKQEHVADNKEVNSNAVAASQGRYTTNDG---VENPADIIEDTGNAYI 202

Db 653 ITTTPETIIOGKPTGDKDPVP-----LANDAFSVLDVYN-----KDFGNASV 695

QY 203 VPHGHHYIIPKSDLSASLAATAKAKHAKNNQPSOLYSTASDNMTQSAVAKSTSKPA 262

Db 696 DANTGIVTTPAKGVESSEPTGTT-----PIIVY-----ODGSVGTIDLAIV 738

QY 263 NKENQSLKELDYDPSAQ-----RYSESDGLVFPDAKISRTFPGVALPHGDHYH 314

```

Db 739 TVSKN-----IYENPGNIPAGYHKTFTAGEGTSTESGTTVAVDGSLPE-----786
315 FIPYKSLALEEKIARMPISGTGSTVSNAR-PNEVYSSLSGSSNSPSSLTSSLSA 373
Db 787 -----DKLPYKAKOG-----YTDKMPPE-----ATOPKADTEFVSSA 822
374 S-DGTYRPNK-----IYETATAYIVRKHGDFHITPKSNQIGOPTL 414
Db 823 TILDIIEPPGNIPAGYHKTFTAGEGTSTESGTTVAVDG-----VSL 868
415 PANSIATPSPSLPIPGTSHKEEDGCGFDANRIIADESEGVMS-----HGH 464
Db 869 PEDKL-----PYLAKADGTTDAKMP-----EATOPKADTEFVSSATKLDIENPGN 919
455 -----NHFEKRD-----LITEO-----IKAAOHL-----486
Db 920 IPAGYHKTFTAGEGTSTESGTTVAVDGSLPEKLPYKAKGTYDAKMPGEATOP 979
487 -----EYKTSHGSLDLSHEDDYGNAKEMKDLKIEKTAGIMKOGVRESITVVK 541
Db 980 KADTEFVSSATKLDKADKADKADKADKADKADKADKADKADKADKADKADKAD 1038
542 E-----KNAIIFPGDHHDADPIDEHKGFGHSHSNYELFKE-EGVAKREGK 590
Db 1039 EKVDSIAGNKGITVYVYSD-GSSDEYE--VDYVTDNRSDADKIEFVEGEKVEIGK 1095
315 VTTGELTNVNNLNKSTFNN-----ONFTLANGORVSESPPEEKKLGIMLVKL 643
Db 1096 VDLTNTVNNLPILPGTITVDTVPGTIDTNPNGVEGIEVETPD-----GTDDVKN 1149
644 -----IIPDKVLEKVGKFGEGV-----GNANFELDOPILPGTFTKTSKYP 691
Db 1150 PVEVTDNRSDADKIEFVEGEKVEIGKVDLTNTNLT-----PLPGT-----1197
315 EYSDGTFTVPSLAKYKASOTIFPFHAGDTYLKVNQFAVPGKTDALVRFDFHNA 751
Db 1198 DTPGCTIDTNPNGVEGIEV-----TY-----PDGKTQVAVPVVETDNR 1239
752 YLENNY-----KV-----GEIKLP-----IPLANGTT-----RTAGNKIPVTMANAYLDN 793
Db 1240 SDADKIEFVEGEKVEIGKVDLTNTNLTPLPGTITVDTVPGTIDTNPNGVEGVI 1299
794 OSTY-----IYVPILEKNOCTKPSILPQFKRKADENSKLDKVEPEKTSKVEK 847
Db 1300 EYVTPDGTQKDTVAVVEVTDNRSDADKTYPMVEGEKVEIGKVD-----1343
848 LSEGNSTNSSTLEEVTVDOVEKAKAFASYGKLENVLFNMGDTIELPLSG--EVI 905
Db 1344 --LIDNTNLTPLPGTITVDTVPG-----GTIDNTNGTEGVELEVTPDGTQKTV 1393
906 K-----KNMADFTGEAPOGNGENKPSENGKVS--TGVENOPTENKPADSLPEAPNEKPV 958
Db 1394 KVPVEVTDNRSDADKIEFVEGE--KVEIGKVDLTNTNLT-----LPQGTITVDTV 1445
959 KPNSTONGKMLNPGN 974
1446 TPGGTIDT-----NTPGN 1458

```

RESULT 5

136868

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st

Species: Plasmodium falciparum

Date: 19-Nov-1988 #sequence, revision 19-Nov-1988 #text, change 09-Jun-2000

Accession: A26868

Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.

Mol. Biol. 195, 373-387, 1987

Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium

Reference number: A26868; MID: 88011243

Molecule type: DNA

Residues: 1-1701 <TAB>

Superfamily: major merozoite surface antigen

C keywords: surface antigen
F/1-19/Domains: signal sequence #status predicted <SIG>
F/20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 4.0%, Score 215.5, DB 2, Length 1701;
Best Local Similarity 20.1%, Pred. No. 0.012;
Matches 249; Conservative 175; Mismatches 407; Indels 405; Gaps 66;

```

QY 1 KFSKXYIAGSAVIVSLCAVALNHOHROEN-----KDNNVSYVDSOSSOSKENTLP 56
Db 500 KMFNNF-----DEVDYDKIFSAITYNVEKORINKNFSSNSVIVY---QKIKALSTLE 553
QY 57 DOVSOKEGIOAEQIYIKITDOGYVTSBGDHYHYNGKVPYDA--LFEEL-----104
Db 554 D-YSLRGI-----SEDFNHYIYLTGKLEADIKLKEIKSSNKILE 596
QY 105 -----LMDPNYOLKADIVN-EYKGYIT-KVDG--KYIYLDAAHADNVT-----149
Db 597 KNFKGLTHSANASLEVSIVKIQVOKVLLIKIEDLRKIELEFKNAQOLKDSIHVPNIYKP 656
QY 150 -----KDEINOKO-----EHVNDKENVSNVA---VASOGRTYNDG---185
Db 657 ONKPEPYLIYLYKKEVDKKEPIPKVMDLKEQAVLSITQPLVASE---TLEDGHS 713
QY 186 -----YFNPADIIEDTGNAYIVPHGHVHYI-----PKSDLSASEIAAKAHIA 230
Db 714 THLSQGETEVEETEETEET-----VGHITVITLIPKRESAPKEVKYENSIE 765
QY 231 GRNNQPSQLSTSSSTRSDNNNTOSVAKGTSKPAK-----SENLSLKEI 275
Db 766 HK-----SNDNSQALTKRTYLYKLDELTKSYCHKIYIVSNSSMOQKLEY 812
QY 276 YD-SPSAORYSESQGLVDPKAKIISRTPNGAIPAIG-----DHYHFIYPSK-----320
Db 813 YNLPPEEKELKS-----CDPLDLFNIONNIPAMTSLYDSANNDLQHLFEELYOKEMITY 868
QY 321 LSALEEK-----IARMPISGTSTVST-NAKPNVYSSLSGSSNSPSSLTSSLSA 373
Db 869 LHKLEENHIKKLEEQOIQITGSTSSPGNTVNTAQASAHNSNQNSNASS--TNT 925
QY 374 SDGYTFNPKDIVEETATAYIVRHGDHFHYIKSNQI--GQPLPN--NSLATPSSSLPINP 430
Db 926 QNG-----VAVSGPVAVESHPDLTVLSISDLKGIIVSLMNGTKVPPN-LTIST 977
QY 431 GTSHEKHEEDGCGFDANRIIADESEGVMSGHDNNHFEKDD-----LITEOIKAAQKH 484
Db 978 -TEMEKTYEN-----ILKNNDTYF-----NDIIOKQVKSNSKYITGLTQKNALNDE 1024
QY 485 LEEVYTS-----HN-----GLDLSHEDDYPGNANEMKD---LKKIEKIKAGIMKQY 530
Db 1025 IKKIKDPLQLSFDLYNNYKLLKDLRFKKKKELGDDKQKQIKLTLTKQLLESKLSLNNPH 1084
QY 531 GYKRE-SIVYNNKKNATII-----YPHGDHHDADIDEKFPVGI 567
Db 1085 NYLONFVFFKKKKEAEIAETENTLENTKILKHKYGLVYKYNGE---SSPLKLTSEVSI 1141
QY 568 --GSHSNYELFKPEEGVAKKEGKVTGEE-----LTNVNLLKNSFNNOPTLA 617
Db 1142 QTEBNYANLEKFRALSITDGLNDNLHGKKKLSFLSSGHLHLETK-EYIKKNKT--1198
QY 618 NGOKRVSPSPPELEKLGIMLVKL--ITPDGVLEKVSCKVFGVGNIANFELDOP 674
Db 1199 -----GNSPSENNK--VNEALKSYENFPPEAKVITVYTP-----PQP 1234
QY 675 YLPGQTYKTYI-----ASKQPEVSYDGTFTVPSLAKYKASOTIFPFHAGDTYLKVN 729
Db 1235 DVPSPSLVSVSVSGSGSTKETQIPTSG-----SLTLEQVVOLONYDEEDSLVYLP 1288
QY 730 QFVAVPGTD-----ALVYFDEFHGNAYLENNYKRGKELKLPKPNKOSTTETAGN 779
Db 1289 IFGESEBNDEYLDQVVTGALISVTMDNITLSG--PENETDVIYIK-PL---AGVYNSLAK 1341

```

[illegible]

QY 417 -----NSLATPSPSLPINPTSH-----EKHEEDGYGFDNRNRIIADES 455
DB 869 EMNEVEYKHAADFVGLGEKHKRAETAKLGEGRREVAVAGLEHAKKE-----VAALLEE 919
QY 456 GFVASHGDHNNHYFPFKDITLDEQIKA--AOKHLEEVKTSYHNSLDITSLSHEDDYPGNAKEMK 513
DB 920 -----KHKEIKALEBGEHKEVMAELGEKKEVYA-----GLEKHNLEBCHKEMVMAELE 968
QY 514 ----DLDKRIIEK-----IAGIMKQYGVKRESIYVNNKNAIIPYHGDDH 554
DB 969 KRAHDVAVLVEEQHKAETIKFGEHKEVYAGIEEKY--KVAAIKLAEBHKDVTYKLGEOH 1026
QY 555 H--ADPIDEKPV-----GIGHSHN-----YELFKPEE 581
DB 1027 KEELAKEDGHEVYVNEVEYKKNASLLNMLENNHNMKIKLEEHKESASDVLVEKLYODE 1086
QY 582 GVAKEGKRYVTGGEELVYVNMILKSTFNNOFTLANQOKVSEFSFPPELEK--KLGIML 640
DB 1087 EV--KSNKKI--ELNIVVIDLINDSIMCYKKOLLEVEKKNENY--EINIKLIYONEM 1140
QY 641 VKLITPDGKLEKVSQ--KVFGEGVGNANFELDOPYLPGOTFFKTLASKDYPEVSYDGF 699
DB 1141 KDM--NDKIKLEKEKEIKLKKLKSNTYKVFETKE-----NTYK----- 1176
QY 700 TVPISLAVKMASQITFFYFHAGDYTLRVNPQFAVPKGTDALY--RVPDFHGNAVLENNY 757
DB 1177 -----NSEMVAVNEKERLITVDSCKENISSEVDS-- 1206
QY 758 KVGEEK--LPIPKINGOTTPTAGNKIPYTFEMANAYLDNQSTYIVEVPLLEKENOTDKPSI 815
DB 1207 KGNMLKMLTSLAKKEERNFTSINDKNNESELVDI--KSAVINKIEYKKEIEDNGANI 1263
QY 816 -----LPQFRKRAQENSKL-----DEVEEPKISEKY----- 844
DB 1264 EDLKNKIIDLSENLINENMKNNVLTDENNNNKKEIEIDNKLNEKKNEMEIILNDI 1323
QY 845 ---KEKSESTGNSNSTNSULEEVPYDPVOEYAKFAESYGKLENYL--FN-----M 891
DB 1324 IKLKEISEMKDDEBKLTKEIN-----KIKNDIEQNNKYKIKKEHMLAKFENINETVSL 1379
QY 892 DGTIELYLPSEGVIAKKNADPTGEAPQNGENKSENGKSVSTGYENOPTENKPADSLPE 951
DB 1380 KNOJIEEMKLEELKNLNEPELLAEKRETNMSISDNDKIVENNILED--TDSK----- 1430
QY 952 APNEKPYAKBENSTDGMLNPGCNVGSDDPMLDPALEAPVDPVOEKLEKFASTAGLGL-- 1009
DB 1431 -----QNNLNK--NVEDKDTGDDINCKRNNDQAKIEISYLDDEIKRISMLYGEELNR 1478
QY 1010 ----DSVIFNMDS--TIELRLPSE 1028
DB 1479 KNSYDEKYNKLTNELKEKLRINKKGE 1504

RESULT 7
A:42404
collagen adhesin - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 04-Mar-1993 #sequence, revision 18-Nov-1994 #text, change 10-Nov-1995
C:Accession: A42404; 527665
J:Patitl, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wlberg, K.; Lindberg, M.; Hook
J. Biol. Chem. 267, 4766-4772, 1992
A:Title: Molecular characterization and expression of a gene encoding a Staphylococcus
A:Reference number: A42404; M01D:32165839
A:Contents: FDA 574
A:Accession: A42404
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 11185 <PAT>
A:Cross-References: EMBL:M01736
A:Note: sequence extracted from NCBI backbone (NCBI:83982)

Query Match 3.8%; Score 205.5; DB 2; Length 1185;
 Best Local Similarity 19.7%; Pred. No. 0.021;
 Matches 214; Conservative 139; Mismatches 411; Indels 323; Gaps 50;

```

36 NNRYSYDGSQSSQSKSENLT--PQVQSKEGIGQEQIYKTTDGGYTSBDDHGHYNGK- 93
b 197 NNEKSYV-----SDITIKDIOGGQOLDISTLININT-----GTHSGYSGOS 240
y 94 --VPYDLFSEBELLKQPNQOLKMDADI-----VNEVKGGIYIKVDS--KYIYYLMDAA 142
b 241 AITDFEKAFFGSKITYDNTKNTIDVTIPGIGSYNSFSINYKTKITNEQKKEFYVNSQAM 300
y 143 HADNVRTKDEINROKOEHVNDNEKVSNNV-----AVASQGRATTN 183
b 301 YQEH--GKEEVNOKSFPHYHTNINANGIGTYKGLKVLKQKDKTAPLANKYFKLSK 358
y 184 DGTYF---NPADIEDTGNAYVPHGCHYHYPKSDLSASELAANAHLAKRMQPSOL 239
b 359 DGSVYKDNQKEIETIDANGI-----ANKALPSGDYIIKEIEAPRYTFDDKE----- 408
y 240 SYSTASDNTQSAKSTSKPANKSENLOSLLKELYDSPSAQRYSDDGLVDPDAKTIIS 299
b 409 -YPTMKOTDNOGYF--TTENAKAIETKDY-----SAQR-----VWGTOYK 450
y 300 RTPNGVALIPGDHYHFIYPSK--LSALEEKIARVPISGSTVSTNAKPNVYVSSLSGL 357
b 451 PTIFYKLYKQDDNQNTPPVOKAEIKKLED-----GTTKVTWMSLTPEND----- 493
y 358 SSNPSLTTSKELSSADSGIYFNPKOIYEETANAYIRHGDHFIYKS--NOJGQPTLP- 415
b 494 -----NSLAPSPSLPI--NPGTSHKEHEEDYQGFANDRIIAEDSGFYVSHGCHNHYF 516
y 416 -----NNSLATPSPSLPI--NPGTSHKEHEEDYQGFANDRIIAEDSGFYVSHGCHNHYF 469
b 517 GYTKKGLVYNTNEKIETTSISGEKVMDDKNDQDKR--PEVYSVNLANG----- 567
y 470 KQDTEQIQAOKHLEEVYTSNGIDSLSSHEQDYGNAKEMKDKLIE-EKIAGIMK 528
b 568 -----EKVYK-----LDVTSETMKY-----EFDLDRKYDEGKRIEYTVT 602
y 529 QYQVK-----RESIYVVK-----EKNALIIYPHGDHHAADIDEKAPGIGHSHSNYELFK 578
b 603 EDHAKDVTDTINGNTTINKYTPGETSATVTKNMDDNNQ--DGKRPETI-----KVELY- 654
y 579 PERGVA-----KREGKRV--YTGELEFVNVLLKNSTPNNOFT 615
b 655 -QQKAFGKATLINESNNHTWTGDEKAKAGQAVKTYVELFV-----KGYTHVDNND 709
y 616 LANGQKRVSESPPELEKKGIMLVKLTIPDGKVLKESGVKVFEG-----VGNIAN 668
b 710 M--GNLIVTKKYPETTSISGEKVMDDKNDQDKRPEKVSNNLADDEKVKTLDTVSETN 767
y 669 FELDQPLP---GQTKYTIAS--KDYPEVSYDGTFTYPTSLATYMAAQTIFYPPHAG 721
b 768 WKYEKFLDPRYDESKKIEYVTEHDVNDY--TTDINGT-----TTNKYTPG 812
y 722 DTYLRV---NPOFVPGTALVVFDE-----FHGNAYLENNYK 759
b 813 ETSATYTKMNDNNQDKRPTKELVYDGAATGATLINESNNHTWTGIDEKAKG 872
y 760 GEIKLPKPKLNQ--GTTTAGKRIPTYFMANAYLDNOSTYIYEVPILEKENDTK--- 812
b 873 QQVYVYEEELTKVGYTHVDNNDGNLIVTKNYPETTSISGEKVMDDKNDQDKRPEK 932
y 813 --PSILPQFRKNA-----QENSKIDE--KVEPETSVEVEKEKSEIGNST 855
b 933 VSVNLLANGKRVTLDTVSETNMYEKDLPKTDGKRIETVTEHDVHYKTYTDTINDTTI 992
y 856 SNSTL--EEVPTVD-----PVOEKYAKFESGKMLENTLFPNNOGYIELYLP 900
b 993 TTKYTPETSATYTKMNDNNQDKRPTKELVYDGAATGATLINESNNHTWTG 1052
y 901 SGEYIKKMAADFT--GEAPDQNGENKRPENGVSTGVENOPTENKRPASL--PEAPNKK- 956

```

Db 1053 LDKRAGQOQKYTYVDELTKVNGYTHVDNNDGNLIVTKNYPETTKPKPKPKDPTP 1112
 QY .957 PVKPENS 963
 Db 1113 PTKPDHS 1119

RESULT 8

SACOM

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)

N:Alternate names: 195K glycoprotein

C:Species: Plasmodium falciparum

C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000

C:Accession: A23386; S06361

R:Weber, J.L.; Leiding, W.M.; Lyon, J.A.

Nucleic Acids Res. 14, 3311-3323, 1986

A:Title: Variation in the gene encoding a major merozoite surface antigen of the huma

A:Reference number: A23386; M01D:86205236

A:Accession: A23386

A:Molecule type: DNA

A:Residues: 1-1104 <WEB1>

A:Cross-references: EMBL:X03831

R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.

Nucleic Acids Res. 16, 1206, 1988

A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria

A:Reference number: S06361; M01D:88143999

A:Accession: S06361

A:Molecule type: DNA

A:Residues: 1104-1726 <WEB2>

A:Cross-references: EMBL:X03831

C:Comment: The merozoite stages of different strains have strain-specific surface ant

C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

F:1-19/Domain: signal sequence #status predicted <Sig>

F:20-1726/Product: major merozoite surface antigen #status predicted <Mat>

F:67-87,91-96,100-109,120/Region: 3-residue repeats (S-G-T)

F:757-765/Region: 3-residue repeats (T-E-E)

F:133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carb

Query Match

3.8%; Score 205.5; DB 1; Length 1726;
 Best Local Similarity 20.0%; Pred. No. 0.038;
 Matches 246; Conservative 172; Mismatches 414; Indels 397; Gaps 65;

```

1 MKTSKTIAGSNVYISLSICAVALNHRSOEN---KONNRYSYDGSQSSQSKSENLT 56
b 531 MKFNNEF---DKQVDKIFESARTYVNEKORYNKSSNNSYVNV---OKLKAUSTYLE 584
y 57 DOVSQKEGIGQEQIYKTTDGGYTSBDDHGHYNGKVPYDA--LESEEL----- 104
b 585 D-YSLRGT-----SEKDPNHYTLTKGLEADIKKTEIKSENNILE 627
y 105 -----LKKDNYOLKDAIVN-EVKGYYI-KYDG--KYIYYLMDAAHADNVRT----- 149
b 628 KNRKGLTHSNAALEVYDIYQVOKVLLIKKIEDLKIELFLKMQKDSIHVPNIYKP 687
y 150 -----KDEINROQ-----EHYKDNKVSNNV---VARSQRYTNDGYVE 188
b 688 QNRPEPYLIVLKEVDKLEFIPKYKMDLKKQOAVLSITQPLVAASE--TTEDG--- 741
y 189 NPADIEDTGNAYVPHGCHYHYPKS-----DLSASELAANAHLAKRMQPSOLSY 241
b 742 -----GSHTHLSQSGTEVEYETETETETGYHTTIVTILPREVKY 784
y 242 SSTA---SONNTQSAKSTSKPANK-----SENLOSLLKELYD--SPSA 281
b 785 VENSIEKHSNDNSOALTKTYLLKLEDEFLKSYCHRYKIIYVSSNSMDQKILEYNTLPEP 844
y 282 QRSSESDGLVDPDAKTIISRPNGVAIPHG-----DHVHFIYPSK-----LSALEEK 327
b 845 ENELKS---CDLDELFIQNIQNNIPMYSLYDSMNDLOHLEFELVQKEKITYYHLKKEE 900

```

```

328 -----IARVPIISGIGSYST--NAKPNVSSLSGSSNPSSTLTSSKELSSASDGIYFN 380
b 901 NHIKKLEEQKQITGISTSSPGNTVYNTAQSATHSNQSQOQSSAAS--TNTQNG----- 953
y 381 PKDIEETATAYIYVRGDDHFIYIPKSNQI--GQPTLPN--NSLATPSPSLDINPTSEHK 437
b 954 ----VAVSSGVAVVESEDPLTVLSTISDGLGIYSLMLGKTKYKPNP--LTIST--TEHEKF 1008
y 438 EEDQYGDANRIIAEDSGFVMSHGDHNYFFKND-----LVEEQIKAQKLEEVKTS 491
b 1009 YEN-----ILKNNDTYF--NNDIKQFVKSNSKYITGLFTQKNALEDEIKKLDFT 1056
y 492 -----HN-----GLDLSHEDDYPGNAKEMD--LDKIEKINGIMKQYVKNR-S 536
b 1057 LQISFDLYNNKTKLIDLFRKKKELGODKQIKKTLKQLESKLSLNNPHAVLNFS 1116
y 537 IYVKEKNATY-----YPHGDHHDADIDEHKPYGI--GHSHS 572
b 1117 VFFPKKKEAEIETENTLENTKILKHYKGLYKYNR--SSPLKLSSEVSIOTEDNYA 1173
y 573 NYELFKPEEGVAKKGNKYVTGEE-----LTNVNLLKNSTFNNOFTLANGQKRV 624
b 1174 NLEKFRVLSKIDGLNMLHGLKKSLSFSLSGHLHLETK--EVIKKNNTY----- 1223
y 625 FSPPELEKIKGIMKVKL--TPDGKRVLEKSGKYPGEGVGNANPELDQYLPQOTF 681
b 1224 GNSPSENNK--VNEAKSYENFLPEAKVTVTP-----PODVTSPSL 1266
y 682 KYTT-----ASKDPEVSYDQTFVPTSLAVKMAQOTIFYPFHAGDTYLRYNPOFAVPG 736
b 1267 SVRVSSGSGSKETQIPTS-----SLTELQVQVQVQVQVQVQVQVQVQVQVQVQV 1320
y 737 TD-----ALVRFVDFEGNAVLENTKVEIKLPKLNQGTTRAGNKIP--V 783
b 1321 NDELDQVVTGEALSTVMDNLSG--FENERDYIYK-PL--AGYRSLKQIEKNIF 1373
y 784 TFMANA-----YLD-----NOSTIYEVPILEKENTDIPSLPQ 818
b 1374 TPNLNDLILNSRLKRRKRYFLDLESOLMOFKHISSEYITE--DSFKILNSEQNTLLKS 1432
y 819 EKRNR-----AOENSKLDEKY-----EKPTEKEREKLETSNSTSNSTLE 861
b 1433 YKTKESVENDIKRQGISYIEKVALYKQDDLESIKYIKKEKEKPPSPPTPPS-- 1489
y 862 EEPVTVDOEKVAKFA-----ESTYKLELVLFM----- 891
b 1490 --PAKTDQKKESEFLPFLNIEFLYNNLVNKKIDYILNLAKINDCNVEKDEARVITK 1547
y 892 -----DGTIELYLPBGGEYAKKMAADFTGEAPGNGENKPSNGK--VSTGVENOP-- 941
b 1548 LSDKAIADKIDLF-----KNHNDFEAIKKLINDTKMDKGLLSTGLVQNFPTI 1599
y 942 -----ENKPADSLPEAPNE--KPYKPN 963
b 1600 ISKLEKRPQDMNISQHCYKQCPENS 1628

```

```

RESULT 9
14859
GA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
Species: Haemophilus influenzae
Variety: strain HK613
Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
Accession: C41859
Poulsen, K.; Reinholdt, J.; Killian, M.
Bacteriol. 174, 2913-2921, 1992
Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
Reference number: A41859; M01D:9234949
Status: preliminary; not compared with conceptual translation
Molecule type: nucleic acid
Residues: 1-1849 <P0D>

```

```

A: Experimental source: strain HK613
A: Note: sequence extracted from NCBI backbone (NCBI:97285)
C: Superfamily: 19A-specific metalloendopeptidase
C: Keywords: hydrolase; metalloproteinase

Query Match      3.8%; Score 203.5; DB 2; Length 1849;
Best Local Similarity 19.3%; Pred. No. 0.054;
Matches 239; Conservative 153; Mismatches 449; Indels 397; Gaps 59;

y 28 HRSQENKNNRVSYDSSQSSQSENTPQVQSEIQ-----AEQYIKITDQ 77
b 125 HRDVSSEENR--YYVEKNFPRENTSFTTEQDAQKREDDYMRDLKFTVEVAPI 181
y 78 GYTSHGDDHYHYNGKPYDALF-----SEELMKDNYOL--KQAD--YNEVKG 125
b 182 EASTANNKGE--YNNSDKYPFLVRLSGSQFYKKSRYOLILEKQKGNLNNMDVGG 240
y 126 YIKVDKYYVYLKDAHADNVRKDEINROKQHVADNEKVNNSVAVARSGRYTTNDG 185
b 241 DNLELVGNATY-----GIAGTPYKNHEN-----NGLIGFNSKEHSDPKG 283
y 186 YV-----FNPADLIEDTGNATYV--PHGGHYIYKSDL-----SASELAALAAHLAK 232
b 284 ILSQDPLNTYAVLIDSGSPLEFYDREKGMFLGSDYFMAQYNKKSQWENNYIYHEPAK 343
y 233 NMQSQSLYS--TASDNNQSV--AKGSTSPARKSENLSLKELYDSPAORY----- 284
b 344 IYQ-----QYSAGSLGSGTQWTQATGTSITTGCGEPLSDLDGDKPRHGSITLKG 399
y 285 -----SESDGLVFPDAKILSRPN-----GVALPHGD--HYHFLPYSKLS 322
b 400 SGRTILNNHIDGAGGLFEEDYEVKGTSDSTWKAGVSAVDKTYWKYHNPKYORLA 459
y 323 -----ALEEK-----IARVPIISGIGSYSTNAK- 346
b 460 KIGKGLVVEGKRGKNEGLKVGDDTVILKQKADANNKVOAQSQVIGSRTSLVNDKQ 519
y 347 --PREVYSSL--GSLSNPSL-----TSSKELSSADGIYFPKD 383
b 520 VDPNSITFGRGGRDLNGSLTDHDIRNIDGARYVNNHNTNTSNITIGESLITNP-- 577
y 384 IVEETATAYIYVRGDDHF-----YIPKSNQIGQPTLPNNSLATPSPSLPIPG 431
b 578 --NTITSYNLEADDDHPLRISIPYQLYFQNDNR--SYTLKKG--ASRSELQNSG 631
y 432 TSHKHEDEGVPD-----ANRIIADESGFVMSHGDHNYFFKNDLTFEQIQAQKHL 485
b 632 ESNENWLYMGRTSDEAKRVNWHINNEKNGF-----NGYF-----GEETKATON-- 677
y 486 EKVTSNGLDLSHEDDY--PGNAKEMDLDKIEKINGIMKQYKRESDIVYK 543
b 678 GKLVWTFNG--KSDNRFLLTGTNLNGDL-----NVEK 709
y 544 NATIYPHGDHHDADIDEHKPYGI-----HSHSNYELFKPEEGVAK-----EGN 589
b 710 GTFLSGRPYRHADI-----AGISSTKDPHFTENNEVYEDDMINNFATIMNTGN 764
y 590 -KYVTGELTNVNNLKNSTFNNOFTLANGQKRVSPPELEKLGISMLVY----- 642
b 765 ASLVSGR--NVANITSNITASN-----NAQVHIGY-----KTGDYCVASDYTGX 807
y 643 LITDGYLVEVSKRVGEV--GNI-----ANFELDQYLPQGFYKTIYAKQDPEVSYD 696
b 808 VTCHNSNLSERKALNSFNPMLNRYNNLTENASFLIGANLFG-----TISIGISQVA-- 860
y 697 GTFVPTSLAVKMAQOTIFYPFHAGDTYLRYNPOFAVPGTDAVRYFDEHGNNA----- 751
b 861 ----LKENSRLHGLGNSVNVQNLNTNGHILHNAQNDANKYTYTTLTVNLSGSGSYTW 916
y 752 -----YLENNRYVGEIKLPI-----PKLNQGT--TTPAGNKIPVTFMANAYL 791
b 917 VDFNNNSNKVYVNRKATGNTLQVADKTGEPNHEMLTFDASVATNNLEVT--LANGSV 975

```

```

Y 792 DN-----GTYIVETILEKENT-----DKPSILPOFKKKAQEN 827
b 976 DRGAMKTKLRVNRKDYLYNPEVEKRNQYDTNTITPNQIQADAP-----AASN 1026
Y 828 SKLDEKVEEP-----KTSEVEKEKISGTENST 855
b 1027 NEELARVETPPPPAPATESALASEQETRPAPTAQAPAMEETNTANSTETAPAKSDTATQ 1086
Y 856 SNSTLEVPYDPOVEAVAPAEISGMKLENYLFNDGTIELYPSGEVTKNNADP-----912
b 1087 ENRPSSEVPS-----ETTEVVAENPPOENETVAKNEDATEPTPQNGEVAKEQPTVEAN 1141
Y 913 --TGEAPQNGENKPSNGSVSTGVENOPT-----ENKPADSLPAPAPKPY-----K 959
b 1142 TQNEATQSSGKNEETQ-----TAEKSEPTSEVTSVSENGPEKTVSGSTEDKVEVEKEK 1196
Y 960 PENSTNGMLNPEGNGSDP-MDPLAEPAPVDPQVE 996
b 1197 AKVETETOKAPQVTSKEPKQAPAEVPTDTMAEE 1234

```

RESULT 10

```

45948
Major mezozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
Species: Plasmodium falciparum
Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000
Accession: A45948
Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
xp. Parasitol. 67, 1-11, 1988
Title: Plasmodium falciparum: gene structure and hydrophathy profile of the major mezo
Reference number: A45948; M0ID:89005525
Accession: A45948
Status: preliminary
Molecule type: DNA
Residues: 1-1726 <GHA>
Cross-references: GB:M37213
Superfamily: major mezozoite surface antigen
Keywords: surface antigen

```

```

Query Match 3.7%; Score 202.5; DB 2; Length 1726;
Best Local Similarity 19.9%; Pred. No. 0.054;
Matches 245; Conservative 172; Mismatches 411; Indels 405; Gaps 65;

Y 1 MKRSKYIAGSAIVISLCAVALNQHROEN-----KONNRSYVDGSSQSKSENLTP 56
b 531 MKRNANP---DKQVVDKISARTTYNEKORNNKTFSSSNNSVYV---OKLKKALSTLE 584
Y 57 DOVSOREGIOAEQIVIKITDOGYVTSBGDHYHYNGKVPYA---LFSSEL-----104
b 585 D-YSLRKG-----SEDPFNHYTLTKTGLEDIKLTLEIKSEKILE 627
Y 105 -----LAKDPNIOKADIVN-EVKGGLII-KVDG--KTYIVLKAADADNVT-----149
b 628 KNEKGLTHSANSLAEVYDYKLOQVKYLLKIKIEDLRKIELFELKKNQOLDSIHVPNIYK 687
Y 150 -----KDEINROK-----EHYKDEKYNANA---VARSGRYTNDGYVF 188
b 688 QNKPEYIYLIVLKEVDKLEKELPKYKDMKKKQAVLSSTOPLVASE---TTEEG---741
Y 189 NPADIEDTGNATVYGHGHYHYTPKS-----DLASELAAKHLAKKNQPSQLSY 241
b 742 -----GSTHTLSQSGEVELEETEETETGHTTYYTITLPPKEVAV 784
Y 242 SSTA---SDNNTQSAKGSSTKRPAN-----SENQSLIKEIYD-SPSA 281
b 785 VENSISKHSNDNSQALTKTYLLKDEFLTSTYCHKYLLVSSNSDQKILEVYNTLTPPE 844
Y 282 QRTSESDGLVTFDAKTIISRTPNCAVAPHG-----DHYHRTPSK-----LSALEK 327
b 845 ENELKS---CDPLDLFTFNQNNIIPAMISLYDSMNDLCHLFFELYOKEMITYLAKLEE 900

```

```

QY 328 -----IARVPISTGTSVST-NAKPEVYSSIGSLSSNPSSILTSKELSSADGYIFM 380
b 901 NHIKILLEKOITGTSISTSPGNTTYVNTASATHSNSQNASAS-----TNNQNG-----953
QY 381 PKDIVEETATATVYRBDHRYIPKSNQI-GQPLPN--NSLATPSPLINGTSEK 437
b 954 ---VAVSSGPAVVEESHDPVLVLSISNDKIVSLNMGKRTYVNP-LTIST-TEKEK 1008
QY 438 EEDGYGDARITIAEDSGVYMSGDNHFFKKD-----LDEQJKAOKHLEEVKTS 491
b 1009 YEN-----ILKNNDTYF---NDDIKQFVNSKRYITGLTETQANALNDEKTKLQDT 1056
QY 492 -----HN-----GLDSLSHEDDYPGNAKEMD---LDKRIEKIAGIKOYGVKE-S 536
b 1057 LQISFDLYNNKYLKLDLRFKKKEKELGODKQKIKLTKLQLEKLSLNNPNVLONF 1116
QY 537 IYVKKENALYI-----YPHGDHHADPIDKRFVGI--GSHS 572
b 1117 VFNKKKEAEIAETENTLTKLKHKGVLVRYNGE---SSPLKLSVSVSIOTEDNYA 1173
QY 573 NYELKPEEGVAKKEGKYVTGEE-----LTNVNLLKSTFNQNTLANQ 620
b 1174 NLEKFRYLSKIDKLDNHLGKRLKSLSSGLHQLTTEKYEKNNKNTGNS-----1226
QY 621 KRVSFSPPELEKKLGINMLVKL---ITPDKYLEKSGVGEVGNIANFELDQYLP 677
b 1227 -----PSENNKK--VNEALKSYENLEPKATYTYTP-----PODPVT 1262
QY 678 GQFFKYTI-----ASQDYPVSYDGFETVPTSLAYKMASQTIFFPHAGDTYLRVNPQA 732
b 1263 PSLPISVYSSGSSSTKEETQIPRSG-----SILTELOQVQLONDDEEDSLVLPF 1316
QY 733 VPRGTD-----ALVAPDEPHGNALLENKYKGEKLIPLKNGTTFATGKIP 782
b 1317 EESDNDVEYLDQVVTGEALISVYKDNILSG--FENEYDVIYK-PL---AAVYSILKQIE 1369
QY 783 -----VTFMANA-----YLD-----NOSTYIVETILEKENTDKPS 814
b 1370 KNFTFTNLNNDLINSKTKRKRFVLDVLESDLMQFHSISENVTIE-DSFKLINSQKNT 1428
QY 815 ILPOFKRNK-----AOENSKLDEK-----EETSEKVEKEKELSTGNTSN 857
b 1429 LKSYKIESVENDIKFAQEGISYEKVLAKYKDDLESIKYIKKEKEKFPSSPTTP 1488
QY 858 STLEVPYDPOVEKAKPA-----ESYGMLE-----NY-----887
b 1489 S-----PVKTEQKKESKFLPLTJNITETLYNVLNKNNDYLLMLKAKKINDCNVEKDEAHV 1543
QY 888 -----LEFMDGTIELYLPSEGVIRKNNADFTGEAPQNGENKPSENGK-VSTGVENG 939
b 1544 KITKSLDKAIDKIDLF-----KNHNDFAIKKILNDYKIMLGLSTGLVQNF 1595
QY 940 PT-----ENKPADSLPAPNE--KPVAPENS 963
b 1596 PNTIISKLEGGKFODMNISQHCYVKKQCPENS 1628

```

RESULT 11

```

A45555
glutamate rich protein - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C:Accession: A45555; S27831
R:Porte, M.B.; Driegel, M.; Hogn, B.; Petersen, E.; Rieneck, K.; Riley, E.; Nels, J.
Mol. Biochem. Parasitol. 49, 119-131, 1991
A:Title: Primary structure and localization of a conserved immunogenic Plasmodium fal
rate life cycle.
A:Reference number: A45555; M0ID:92131041
A:Accession: A45555
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1271 <BOR>

```


A:Cross-references: EMBL:M59706; NID:g160311; PID:g160312
 A>Note: sequence extracted from NCBI backbone (NCBIN:77801, NCBIPI:77802)

Query Match 3.6%; Score 196.5; DB 2; Length 1271;

Best Local Similarity 19.1%; Pred. No. 0.067; Mismatches 434; Indels 333; Gaps 55;

19 SLCAVALNHRSEKNDNRNRYVYDSSQSSQSENLTPQVSKQEGIQAEQIYIKITDGS 78
 153 SLEDELLENSSDNDLKDPTIS--TEPPNOKKDLQOD-----LNDEPL-----EP 197
 79 YVSHGDHYHYNGKVPYDALFSEELMKDPNTOLKADIVNEVKGYYLIKVDGKYYVL 138
 198 FPPQIKHDKYEKN-----LINEDESEPPROKHKVDNHNNEKN-----VFH 239
 139 KDAAHADN--VTKDEINROKQEHVKNKNSNAVAAR-----SGRGYTN-DGYV 187
 240 ENGSANGNOSILKSPDEHLKDEKTEENELVHENLSTIPNDPIEQILNDEQETNIOEL 299
 188 FNPADIEDTGMAI-----YVHGCHYHYTPKSDLSASGLAAKAAHLAGKNA 234
 300 YNEKONVEEKNOSQISLIDKEPTNEDILNHNPLENIKQSE--SEINHYODHALPKEN 356
 235 QPSQSYSSASDNNQSVAKGSKTPANKSENLOSILKELYDPSAQRYSSESDGLVFPD 294
 357 TIDL-----DNQEHIDQSOHNINVLQENNINNHQLEPOEKPNIES-----FEP 401
 295 AKIISR--TPNGVA-----IP--HGDIHYFIPYSKLSALEEKTARAWPISGTGTVS 342
 402 KNIDSEILLPENVEETIIDVSPKSNHETFEETSESEHEEVAVE-----KNAH 453
 343 TNAPNVEVSSLSLSSNPSLTTSKLSASGCIYTPNDIYEBATAY-----IV 394
 454 EYVEHEETVSQ--ESNPEKADNDGVNSQNSNNEL--NENEFVESEKSEHEAENESSL 508
 395 RHGDHFIYIKSN--QIGQPTLPNNSLATPSPSLIPNGTSHKHEHD-----GYGFA 446
 509 EECHEEIVEQNNNEESGKLYDN-----DEGFEENHNESESSEVSNSELINE 557
 447 NRIIADESGFVASHGDHNYFFPKDLTE--EQIKAAKHLEEVTSINGDLSLS-- 500
 558 NEVEBSKS--VTEPAHEHYVSESENPAPAESESSIEAHQELIYEPENDESSESSL 615
 501 -----HEODYPG-----NAKEKDKLKEEIKAGI--MKQYGVKRESIV 538
 616 VDNREGDEEERNHEEFPDNDSELSENEIYSEKSYSEPAHEVEIYSEKSVSEPAHEVE 675
 539 VNEKNAIIPRODHHDHADDIDENKPYGCHISNVELEFREPESVAKKEG--NKVYTG 595
 676 IYSEKST-----SEPAHEVES--EOSNNE--PSE--KKDPPVDSKPEET 716
 596 ELTNVVMILKSTNNQNF--TLANGOKRVYSFSP--PPELEKKGILNMLVLTTPD 647
 717 EKYDVQPKIYDIDILIBENFDSQNPQEPYEPSPVKEKYPSEKNKASV-----D 767
 648 GAVLEKYSGVFEGGYNANFELDQYLPDQTFKTTIASKDYPERYSYDGTFFVPPSLAY 707
 768 PEVKEK-----ENVSVEVEK-----QNSQSEVEEIPVNEDEFEFVHTEQDLIDH 812
 708 KMASQTFYFPHAGDYLVRNPOFAPVKGTDALVRVDEFGHNAUYLENNKVGKILPIP 767
 813 K-----YDPEIV-----EVEEIPSEIHEEVA--HEIYIEIEVEYF 847
 768 KLNQGTTRAGNKIPYFMANAYLNDQSY--IVEV--PILEKNGQTDK-----PS 814
 848 EPNQNNEFOGINE-----DKSAHIQHEIYEEVEILPEDDKNEVEHEIYEEVE 896
 815 ILPOFKRNKAQ--ENSKIDVEEPEKTSKYE-----KEKLSGTGNSSTLEEVTVD 867
 897 ILPEDKNEKQOHEIYEEVEILPEDDKNEKYEHEIYEEVEILPEDKNEKQOHEIYEEVEIL 956
 868 PVQEKVAKFASVGMKLENTL--FNNDG-----TILLYLP--SGEVIKKNMADFTGEAP 917

DB 957 P-EDKNEK-VEHEIYEEVEILPEDKNEKQOHEIYEEVEILPEDKNEKQOHEIYEEVEIL 1014
 QY 918 OCGENKPSNGKSVSTGYENOPTENK-----PADS----- 948
 DB 1015 EDKNEKQOHEIYEEVE--EILPEDKNEKQOHEIYEEVEILPEDKNEKQOHEIYEEVEI 1070
 QY 949 LPEAPNKE-----PVKPNSTDNQMLPBGVNSGDPMLDPALEE 987
 DB 1071 LPEDKNEKQOHEIYEEVEILPELVEIEVPSQNNNENETIKPEEKNEFSVEEKAIPQ 1130
 QY 988 APAPDPVQE 996
 DB 1131 EPVPTLINE 1139

RESULT 12

F82884 hypothetical protein U0495 [imported] - Ureaplasma urealyticum

C.Species: Ureaplasma urealyticum

C.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C.Accession: F82884

R/Glaas, J.I.; Lefkowitz, E.J.; Glaas, J.S.; Helner, C.R.; Chen, E.Y.; Cassell, G.H. submitted to GenBank, February 2000

A.Description: The complete sequence of Ureaplasma urealyticum. Alternate views of

A.Reference number: A82870

A.Accession: F82884

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-5005 <GLA>

A:Cross-references: GB:A002148; GB:AF222894; NID:g6899495; PIDN:AAF30907.1; GSPDB:GN

A.Experimental source: serovar 3; biovar 1

A.Gene: U0495

A.Genetic code: SGC3

Query Match 3.6%; Score 196; DB 2; Length 5005;

Best Local Similarity 17.3%; Pred. No. 0.59; Mismatches 454; Indels 490; Gaps 57;

Matches 234; Conservative 177; Mismatches 454; Indels 490; Gaps 57;

QY 6 KYIAASAVIVS-----LSLCAVALNHRSEKNDNRNRYVYDSSQSSQSENLTPD 57
 DB 38 KNYEGEDVYVNVQDVGNLIKIPALYNNKHQIDLSNN-----PMTKD 83
 QY 58 QVSQKEGIAEQIYIKITDGYTSHGDIHYNGKVPYDALPSE----- 103
 DB 84 KIYHFDISIE-----DEYH-----IPKETIISDDVDNSVTLNNEVD 122
 QY 104 -----LAKDQNV-YOLKADIV-----NE-VKGGYIIVDGYVY----- 136
 DB 123 LGRKITYKNPSVIONKPTTILTVNDPNKIALNERKLTITLALNDRPKHCVSAGLIDYKN 182
 QY 137 -----YLKDAHADNVRTDEINROKQEHVKNKNSNAVAARSGRY 180
 DB 183 NSLVEFENHLOTFAVYIKLVKNNNLNLTDLNAAHSGQTKIN--LIKQYIGYFRIF 240
 QY 181 TINDGYFNPADIIEDTGNAIYVHGCHYHYTPKSDLSASGLAAKAAHLAGKMOQSLS 240
 DB 241 STN-----TLVLED-----YHOY-DTPNARKL--NLNISTTKHRTIT 277
 QY 241 YSSTASDNNNT-----QSVAKGSTSKPANKSENLOSILKELYDPSAQ-RYSE-- 286
 DB 278 YLLDENNALPELDVLANLYLVKKNYVAHLONLKGNSKTYLLGFRDENNNKIHSTPI 337
 QY 287 SDGLVF-----DPAKISRTPNGVALIPGDIHYEIPYS--KLSALEEKIA 329
 DB 338 KTFILIGTTPHATIVFVNVNPKOITKSQARAIIDPVDDQKLRIGNPYELTYLKNKNQ 397
 QY 330 RAVPISGTGTVSTNAKPNEVSSLSL--SSNSSLTTSKLSASADG--IFNKD-- 383
 DB 398 QYKTIQYVGVINKTKNSGTLPELNLNLSNNNYELLEKALGLANDAYOKOVFAFPDOK 457

```

2Y 384 -----IYEATATAYVHAGDH-----EYIPKSNQI 409
Db 458 HNOFSTKLGITYLESISSEVYIKKDKNDLVNKKIYFSQDKTLNNVMAQLYISKDNV 517
2Y 410 --GQP-----
Db 518 CLSQPLLLLVGNSNTYFTLTNNKPNQFSLNRYKKFVLLIYADQDNDLQSTDAVKNSTDL 577
2Y 413 -----TLPNNSLATPSP-----SLPINSPTSHKEHEEDGYGF-----D 445
Db 578 SHGKTEPTGIISLKNNSLQVYIKRTDSDIRLSIPL-VQNDHVLVYDSDGSFRTKISDLD 636
2Y 446 ARIILDEDSGFVMSG-----DHNYFPKKDLTEQIKAA--QKHLVEY 488
Db 637 PNNLPDADQKVIYDQNGRNMLEFDPVNSLNNKHYNINIIITLKNPINAALFQNEHQV 696
2Y 489 KTSNGLDLSLSS-----HEODYPGNAKEMD--LDKRIE--EKIADIMQYGRSS 536
Db 697 YELNMQPOLSEFQNELLYNTYNTNPTDVTNLTNLYKVLTKSLMLMEYIKLY 756
2Y 537 IYVNEKNAIYPHGDHHAADPIDEKPYGIGHSHSNYELFPDEGVAKKGNKYTGEE 596
Db 757 IDNN-----GDVMSD-----PVSINNVTNVEFNNLPDKALK--SNRIYKFG 798
2Y 597 L-----TNYNLIKSTFNQNPFTLANGQKRVSPSPPELEKLGIMLVKLTDPGKVL 651
Db 799 LYFQKDVNTVKITSARKVMMNIT-----PLOCQASKINLNSPVT----- 840
2Y 652 EKVSQVFEQGVGNINAFELDQPLP-----GQFFKTYLAKQDPVSYQSTFTV 701
Db 841 -----INNTIPASADL-YLPVSTSDIIPAKQDVIQVITGNKD--NKNQDMFT 885
2Y 702 PSLAYKASQITFPFHADYTLRVN-----
Db 886 -SNLEYDDITKTGWGVYIHNNTLSPEVNYQIKSVKFRQKQVLAFAVNNSDHVLDSQK 944
2Y 729 -FOFAPKKTDLVYFDFEFGNAILIENYKVEIKLPKLNQGT----- 774
Db 945 TGFSTPKASFPLVSV-----TASDVYDANKVKNSHAIYIN--NDGSLNCRKAKIYVNG 998
2Y 775 -----RTAGNKIPYTFMANAY-----LDNQSTIYVEPILEKENQD--KPSILPQFRNKA 824
Db 999 EHDLYSNFVYLVANGVNEDEFLMDLKHNRITSEKLLITEPNQMDYFSFL----- 1050
2Y 825 QENSKLDEKVEEPTSEKVEKE-KLSEGNSTSNSTLEEVPTVDVQOEKAFASGYK 883
Db 1051 -NNOKITWTEKASVNEQITIDFKLVKPKDSMLNKLQINDP----- 1094
2Y 884 LENVLFNMDGITELLPSEGVYKKMAADFTGAPQNGENKPSSENGKYSTGTVENOPTEN 943
Db 1095 --NDPLEDSDILEITFHDEK--DKTTHVYIGKINVDANNKLTLEFSVENTNFKIOPNKK 1150
2Y 944 KPADSLEAPNKKPKPENSNDNGMLNBEANGVSPMDPLALEAPDAVDPOEKL----- 998
Db 1151 YVVDVINTATKKK--IQPANAISN--NSNKSITDASSNPS-----KILLSFTN 1194
2Y 999 EKFTASYGIDSVIFNMDGITELLPSEGVYKKN 1033
Db 1195 ELFVNININVPNTMLNPTASIDVELKSSQNLMD 1229

```

```

RESULT 13
18447
Hypochemical protein C0335c - malaria parasite (Plasmodium falciparum)
Species: Plasmodium falciparum
Date: 15-Oct-1989 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
Accession: 118427
Lawson, D.; Bowman, S.; Barrell, B.
Submitted to the EMBL Data Library August 1997
Reference number: 218935
Accession: 118427
Status: preliminary; translated from GB/EMBL/DBD
Molecule type: DNA

```

A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:298547; NID:el325376; PID:el325379; PIDN:CAB11104.1
C:Genetics:
A:Introns: 307/1; 1545/2
A:Note: C0335c

Query Match 3.6%; Score 193; DB 2; Length 3724;
Best Local Similarity 16.7%; Pred. No. 0.52;
Matches 185; Conservative 185; Mismatches 381; Indels 358; Gaps 49;

```

QY 30 SOENKNNRSTVDSQ--SSQSENFDPQ-----VSQREGIQ-----AEQIVKITDQ 78
Db 446 NNEENIDNN--IYVENNKLKQDSYDLFSESEGNLILGVNEGEFNEVEFENIEKLEQER 503
QY 79 YVTSQDHVHYNGVPPDALFSEELMKDPYQKADIVNEVGGIILVQKYYIYL 138
Db 504 DEKNNDKTYINNNEQTD-----DLNRNINKTESIN----- 535
QY 139 KDAADAVRTDEINROKOEHVKNENYNSVAVASQGR-----YTTNDGYFNPADI 193
Db 536 NNDNNNNNNKKEFNKIRTEHILNKESIKHIGSPSRDKKIKLYTNK-----NEDST 591
QY 194 IEDQNAVIVPHGHIYHIPSDLSASELAAKAHLAGKNQPSQSYSTASDNNTOV 253
Db 592 FELKKELELITNNKVNYY--EEDIIGSNEDEYIHLKENLEKEDANEYN----- 638
QY 254 AKGSTKSPANKSENQSLKELDYDPSAQRYESDGLVDPKIIISFTPNQVAPPHGHY 313
Db 639 -----NDKENNNKTKELKS--KNILENKRILEELK--KQKNNT----- 676
QY 314 HETPYKLSALEEKIARVWPISGTSTVSTNAKPNEVYSSLSGSSNSSSLTSKELSSA 373
Db 677 -----FKQEKYNSLGEVILNIGQINEKKINDI 705
QY 374 SDGYFNPROVIEPTATYIYVHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINCTS 433
Db 706 QDGNISKOR-IIQSSS-----RTNDTFNI-----KDISLNDLE 738
QY 434 HEKHEEDGYDANRIILDEDSGFVMSGDHNNHYFFKDLTEQIKAAQKHLVEY----- 488
Db 739 KEKRRKKSQHFIDNKVD-----KNEISENINKICDNNINNIYDESI 781
QY 489 -----KTSNGLDLSLSEHD-----YPNAKAKED--LDKLTIEKTAGIMKOYGVAR- 534
Db 782 NNIDSEINNIYDESINNIYDESINNIYDENINNIYDENINNIYDENINNIYDE-GIKKI 840
QY 535 -ESTIVNK-----EKNAIYPHGDHHAADPIDEKPYGIGHSHSNY 574
Db 841 CDDNILENKNIKTTNDIYQVEENNESIEKNEML-----ISLNKDI-----NNTY 884
QY 575 ELFK-----PEEVAK--KEGN-----KYVTGE-----ELTN----- 599
Db 885 NMFKENVDIFINKIRRESLAKIDKNINDNNDDEYINDNFEENFIINHEKKEITNKLEDP 944
QY 600 -----VNL-LKNSFNQNPFTLANGQKRVSPSPPELEKLGIMLVKLTDPGK 649
Db 945 LEITQONEFINLDIKKKYIYNDF--FNDADKAFIEAKKILNDAKKNEQEFKFD-- 1000
QY 650 VLEKVSQVFEQGVGNINAFELDQPLPGQTFKTYLISKDYPEVSYDQTFVPTSLAYKM 709
Db 1001 -----ETFGLSQSHKI-KYNNKGE-KHDKNNEEKNILYD----- 1034
QY 710 ASQITFPFHAGDYLVKNPQFAPVKGTDALVRFDEHG--NAYLENNKYVEIKLP 766
Db 1035 -----ENQYTSVLYSDHKIEDQIDQIHISIQITICDEN-----I 1068
QY 767 PKLNQGTTRTGNKIPYTFMANAY-LDNQSTIYVEPILEKENQDPSILPQFRNKAQ 825
Db 1069 EQIENENSK-KGVNISGTDEKNDAENKND-----MEKKNDMEKKNDIE--KKNDME 1118
QY 826 ENSKLDKVEEPTSEKVEKEKLSEGNSTSNSTLEEVPTVDVQOEKAFASGYKLE 885

```



```

;Reference number: 219322
;Accession: T20774
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 14667 <112>
;Cross-references: EMBL:281499; PDB:CA54224.1; GSPDB:GN00019; CESP:ZK270.2d
;Experimental source: clone F11C3
;Kerishaw, J.
Submitted to the EMBL Data Library, November 1996
;Reference number: 220423
;Accession: T27818
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 14667 <112>
;Cross-references: EMBL:282089; PDB:CA54513.1; GSPDB:GN00019; CESP:ZK270.2d
;Experimental source: clone ZK270
;Genetics:
;Gene: CESP:ZK270.2d
;Map position: 1
;Intons: 61/3; 112/1; 244/3; 328/2; 384/1; 406/3; 4302/3; 4343/3; 4384/1; 4422/3; 4461

Query Match          3.58; Score 190.5; DB 2; Length 4667;
Best Local Similarity 17.18; Pred. No. 0.99; Mismatches 448; Indels 443; Gaps 53;
Matches 219; Conservative 167;

Y 57 DOVSQKEGICQEQIVIKITDQGYTSHGDHYHNGKVPYDALFSEELMDPNYQLDA 116
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2551 DEISRETDLDH----APIADANANYHGGDSRIEKHASPATLEKKOKKVASPKKEKVDY 2606
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 117 DIVNEVGKGIYIKYDGKTYTYLKDAADNRYTDE-----INROKEQVKNDEK--- 166
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2607 PAEPKEKIRIJA-----RVTEPEAEIYDVVPVKKETKEKKGSKSLTP 2650
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 167 ----VNSVAVASGGRATYNDGYV--FNPAIDIED--TGNAYIVPHGGHYIYIKSDLSA 219
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2651 TFGPFSKTKTKTKTKTGPEISEAYTGETDTRIHDLGEGTSF--EHGHPAYSIRKQYIE 2708
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 220 S-----ELAAKAKILAGKNM--QPSQLSYSTASDNNTOYAK----- 255
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2709 ALEPPEPEKRYRLIAFRHEGDEVDYKPDYQIATLYLDGPLDEISRDTDHPPIA 2768
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 256 -----GSTSK-----PANKSENQSLKELIYDPSAKRYSSESGLVFPDPAKI 297
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2769 DHANVYHGGDSRIEKHASPATPEKKOKKVASPKKEKVDVPAPPEKKE-----IRL 2820
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 298 ISRTPNGVAIIGHGHYHPIPSKLSALEEKIARAVPISGTSVSTAKPNEVYVSLGSL 357
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2821 IARVRY-----EPPEAEIYDVVPVKKETKEKKGDSKSLPTFGFF 2858
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 358 SSNPSLSLTS--KELSSASDGIYFNPKDIYETATAYIVRHGDHFHYIPKSNQIGOPTLP 415
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2859 SKTKTKTKITGYPEISEAYTGETDTRIHDLGEGTSF--EHGHPAYSIRKQYIE 2910
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 416 NNSLATTG--PGLPINPGTSHKHEEDG-----YGFDAKRITIAEDSGFVM 459
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2911 -ETIETPAPEPKKRYRLIARRHGDEVDYKPDYQIATLYLDGPLDEISRDTD-----L 2965
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 460 SH---GDHNHFFKKDLTEQIKAAOKHLEVKTSHGGLDLSLSHEDYDPNAKEMDLD 516
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 2966 DHTPIADANANYHGGDSR-----TEKASPVAT-----LEKKOKKVASPKKEKVDVP 3013
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 517 KRIEKTAGIKQYGVKRESIVN-----KEK-----NAILY 548
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 3014 AEPKEKELRIARVTEPEAEIYDVVPVKKETKEKKGDSKSLPTFGPFSKTKTKTKTG 3073
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 549 PPGHHAADPID---EKKPVGIGSHSNYELFPEEGVA----- 584
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 3074 PETSEAYTGETDTRIHDLGEGTSFEGHHPAYSIRKQYIEATETPAPEPKKRYRLIARF 3133
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Y 585 KKEGKN-----YLTG--EELTANVVLKNSFTNNQNF----- 614
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
b 3134 RHGDEVDYKPDYQIATLYLDGPLDEISRDTDHTPIADANANYHGGDSRIEKHAS 3193
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

QY 615 --TLANGQRV-----SFSPELEKKILGINLVLLIT--PDGKYLEKVSQVYEGCGVG 664
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3194 PVATLEKKOKKVASPKKEKVDVPAPPEKKE--EIRLIARVTEPEAEIYDVVPVKKETKEKK 3252
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 665 NIANFELDQYLLPQGTQK-----TIAKDQYEVVS--YQG-----PFTVP 702
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3253 D-----SKSKLP--TGFPSKTKTKTKTKTGYPETSAVYTGELDTDRIDLEGTSEHG 3304
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 703 TSLAYKMASQTI-----FYPP-----HAGDYTLRNPQFAVPKGTDAL 740
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3305 EHPAFEASSITITLEADPDSHEKVIKPSRFHMSRRRHGDEVDYKPDYQIATLYLDGPL 3357
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 741 VRYVDEPHGNAYLENNYKVGELKIPRLNQGTTTIAKNIKLVETEMANAYLDNOSTIYVE 800
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3358 --LSSEIYEGPY--DNMAPIGLEL-----TPILDHQVYHNGSGKLEK 3398
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 801 VPILKENQDTP-----SILPQFKRKAGQNSKLDKVEEPTSEKV----- 843
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3399 RSSKEKIKTPSPERVPSEIKLVARIQPLEHVEDSISEKLSPVKDRSRAPFSMFHRQ 3458
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 844 EKER-----LSEIGNS--TGNSTLE--EVPYDVPOVEKVAKFAESYGMKLENY 887
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3459 SKQGYPEISPIYEGNIDYVGRALVEMENTPIEVASVPEIYHPKQEVSAQHES--SQPNR 3515
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 888 LFNKDGTEIYLPSEGVYIKKNMADFTGEADQG-----NGEN 923
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3516 FYRLIARRHGDEDELEKPDAYKFTVEYEGPLDIPVVEIADQPISDHSQVYHNGES 3575
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 924 -----KPSENGK-----VSGTGVNQPTEKPKADSLPEAP 953
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3576 WKAEKNIKPPRPVANTKKGISPEIESEVHEIRLLITVNASSEPEEVDYIAPASTVYKP 3635
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 954 -----NEKPYKPNSTNDGMLNPEGNY-----GSDPMLDPALEEA 988
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3636 FRLRFLSIGKSPSPSEPEECSPQYDGPVLDLSRSDLEVMYLLGVSVPYSPKFTPV 3695
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 989 PAVDPVOEKLKFTASY 1005
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3696 KKYDKADEERKKYVIY 3712
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Search completed: September 26, 2001, 22:32:23
Job time: 307 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

3M protein - protein search, using sw model

Run on: September 26, 2001, 21:37:01 ; Search time 57.46 Seconds
(without alignments)
1096 212 Million cell updates/sec

Title: us-09-471-255-2

Effect score: 5406
Gene: 1 MKESKRTIAGSANTVLSLST.....TELRLPGEVYKRLSPFIA 1039

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

1: /SIDSL/gcgdata/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5406	100.0	1039	21	AA12715 Streptococcus pneu
2	5315	98.3	1019	21	AA12722 Streptococcus pneu
3	5315	98.3	1019	21	AA12751 Streptococcus pneu
4	5297	98.0	1019	21	AA12750 Streptococcus pneu
5	5297	97.9	1019	21	AA12748 Streptococcus pneu
6	5291	97.9	1019	21	AA12749 Streptococcus pneu
7	5291	97.9	1019	21	AA12752 Streptococcus pneu
8	5291	97.9	1019	21	AA12753 Streptococcus pneu
9	4276.5	79.1	840	21	AA12721 Streptococcus pneu
10	4090	75.7	780	21	AA12744 Streptococcus pneu
11	2953	54.6	568	21	AA12731 Streptococcus pneu

12	2953	54.6	1057	21	AA12725
13	2746	50.8	509	21	AA12719
14	2650	49.0	509	21	AA12724
15	2559	47.3	489	21	AA12723
16	2494.5	46.1	485	21	AA181538
17	2492	46.1	484	21	AA181467
18	2492	46.1	484	21	AA181708
19	2485	46.0	484	21	AA12718
20	2338	43.2	447	19	AA161228
21	2128	32.0	329	21	AA12732
22	1307	24.2	840	21	AA12716
23	1295.5	24.0	819	21	AA181468
24	1286	23.8	819	21	AA181469
25	1286	23.8	819	21	AA12720
26	1282.5	23.7	826	21	AA191939
27	1280.5	23.7	838	21	AA181466
28	1277.5	23.6	827	21	AA181662
29	1266.5	23.4	821	21	AA12727
30	1266.5	23.4	821	21	AA12766
31	1264.5	23.4	690	21	AA12745
32	1263.5	23.4	811	21	AA12763
33	1263.5	23.4	834	21	AA12759
34	1249.5	23.1	796	19	AA181765
35	1247	23.1	796	19	AA181765
36	1243.5	23.0	811	21	AA12760
37	1243.5	23.0	811	21	AA12762
38	1240.5	22.9	816	21	AA12757
39	1239.5	22.9	816	21	AA12756
40	1238.5	22.9	819	21	AA12740
41	1238.5	22.9	819	21	AA12754
42	1237.5	22.9	805	21	AA12764
43	1236.5	22.9	811	21	AA12761
44	1233	22.8	820	21	AA12755
45	1231.5	22.8	816	21	AA12758

ALIGNMENTS

RESULT 1
ID AA12715 standard; Protein; 1039 AA.
XX
AC AA12715;
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-3 protein antigen SEQ ID NO:2.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-26; antigen; vaccine;
KW Prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
(BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX MPI: 2000-452397/39.
XX N-PSDB; AAA65730.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteremia and/or pneumonia

35 Claim 18; Fig 2; 106pp; English.

36 The present invention describes nucleic acids (I) encoding protein
37 antigens (II) from Streptococcus pneumoniae. The protein antigens
38 have bactericidal activity. The nucleic acids, encoding the protein
39 antigens, may be used for the recombinant production of the proteins
40 they encode. The protein antigens may then be used as vaccines for the
41 prevention and treatment of Streptococcal infections in mammals
42 (especially humans) which result in, e.g. meningitis, otitis media,
43 bacteraemia and/or pneumonia. The present sequence represents the
44 S. pneumoniae BVH-3 protein antigen.

45 Sequence 1039 AA;

Query Match 100.0%; Score 5406; DB 21; Length 1039;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1039; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKFSKRYIAAGSAVIVSLCAVALNQHROENKDNMRVSYVDSOSSKSENLTPOVY 60
1 mkfskryiaagsavivslcavalnqhroenkdnmrsvyvdsgssksenltpdqvs 60
61 QKEGIAQDIYKIDGQGYVSHGSHYHYNKVYDALFSEELAKPQNLKADIVN 120
61 qkegiaqdiiykidggyvshgshyhyngkvydalifseellmkpnyglkadivn 120
121 EVKGGYIIVDKYVYVYKDAHADNVRTKDEINRQKEHVKNKNSNVAVRSQRY 180
121 evkgyyilkvdgyvyykdaahadvrtkdeingrkehndknsnvavarsqrry 180
181 TTNDCGYVNPADIIEDTGNAYVTPHGSHYHYTPKSDLSASEIAAKAHLAGKNPQSL 240
181 ttndcgyvnpadiiedtgnaylvphgshyhytpksdlsaseiaaakahlagknmpsqsl 240
241 YSSTASDNNTOSVAGSRSKPNKSENLOSLKELYDPSAQRSQSESGIVPDPKTIIS 300
241 ysstasdnntosvagsrskpnksenloslkeltydpsaqrsesgsvpdpkttisr 300
301 TPNGVAFPHGSHYHYTPSKLSALEEKIARWPIGSGSTVSTNAKPNEVVSLSGLSSN 360
301 tpngvafphgshyhytpsklsaleekiarwpiigsstvtstnakkpnevvsllsglssn 360
361 PSSLTSELSASAGYIFPNKDIYEETATAYVARGHSHYHYTPKSNIGOPTLPNSLA 420
361 pssltseelsasagytfnpkdiyeetataiyvarghshyhytpksnigoptlpnsla 420
421 TPSPSLPINPGTSHKHEHDGYPGPNARLIIAEDSGFYMSGDHNYEFKKDLTEQITKA 480
421 tpspslpinpgtshkhehdgygpnarlliaedsgfymsgdhnyeffkdlteqitka 480
481 AOKHLEEVKTSHGSLDLSHSEODYPGNAKEMKDLKKEIKIAGIMQYGVKRESIYVN 540
481 aokhleevktsngslslshseodypgnakemkdlkkeikiiagimqygvkresiyvn 540
541 KEKNAITTPHGDHHDADIDDKRPVYGISHSNTLEKPEEGVAKKEGNYTGBELTNV 600
541 keknaityphgdhhdadiddkrpvygishsntlekppegvakkegnytgbeletnv 600
601 VNLKNSFNNOFTLIANGOKRVSFPPELEKLGIMTKLITPOKTLVEXYSGKRYG 660
601 vnlknsfnnoftliangokrvsfppeleklgimtklitypokitlvexysgkryg 660
661 EGVGNIANFELDQPTLPQOTPKYTIASKNDEPVSYDGTFTVPTSLAYMASQITFPFHA 720
661 egvgnianfeldqptlpqotpktytiaskndepvsydgftftvptslaymasqitfppha 720
721 GDTYLRNVPQAVNFKGTALVAVPDERHGNAYLENNIVGKIKLPIPLNCGTTRTAGNK 780
721 gdtylrnvpqavnfkgtalvavpderhgnaylennivgkiklpiplncgttrtagnk 780
781 IPVFMANAVYLDNOSTYVVEPILEKENQTKPSILPOFKRNKAQENSLDEKYEERTS 840
781 ipvfmanavyldnostyvvepilekenqtkpsilpofkrnkaqensldekyeerts 840

Db 781 lptfmanayldngstylvvepilekenqtkpsilpofkrnkaqensldekyeerts 840
QY 841 EKVEKEKLETSNGSTNSNLTLEVPVPOEERKAVFAESYGMKLENNVLFNNDGTIELTP 900
QY 841 ekvekekletsgnstnsnltlevpvpoeerkvafaesygmklennvlfnnndgtieltp 900
Db 841 ekvekekletsgnstnsnltlevpvpoeerkvafaesygmklennvlfnnndgtieltp 900
QY 901 SGVYIKKMAADFTGAPQNGENKPSKNGKYSTGTVENQPIENKPADSLPAPNEKPYKP 960
QY 901 sgvyikkmaadftgapngenkpskngkystgtvenqpienkpadslpapnekpypk 960
Db 901 sgvyikkmaadftgapngenkpskngkystgtvenqpienkpadslpapnekpypk 960
QY 961 ENSDNGMLNPGNVGSPMDPALAEAPVPOEKTEKFTASVGLDGVIFNMOSTI 1020
QY 961 ensdngmlnpgnvgsppmdpalaeapvpoekektekftasvglgdvifnmosti 1020
Db 961 ensdngmlnpgnvgsppmdpalaeapvpoekektekftasvglgdvifnmosti 1020
QY 1021 ELRLPSGEYIKKNSDFIA 1039
QY 1021 elrlpsgeyikknsdfia 1039
Db 1021 elrlpsgeyikknsdfia 1039

RESULT 2

AA012722
ID AAB12722 standard; Protein: 1019 AA.

AC AAB12722;
XX 21-NOV-2000 (first entry)
DE Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

XX MO2000039299-A2.

XX 06-JUL-2000.

XX 20-DEC-1999; 99MO-CA01218.

XX 23-DEC-1998; 98US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Plneau I, Martin D, Rioux C, Charland N;

XX WPI: 2000-452397/39.

DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
PT Claim 18; Fig 20; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3M protein antigen.

XX Sequence 1019 AA;

Query Match 98.3%; Score 5315; DB 21; Length 1019;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CAYALNHRQENKDNMRVSYVDSOSSKSENLTPOVQKQIAQVIVITTOGYV 80
QY 21 cayalnhrqenkdnnrvsvydsossksenltprovqkqiaqvivittoogyv 80


```

> 1 cayalnhqrsgeknknrvsyvdsqsgsksealtpdqvsqkqegiaeqivlkltdggy 60
> 81 TSHGCHYHYNKNKVPYDALFSEELMKDPNYOLKADLYNEKGGYIIKVGCKYVYVYKLD 140
> 61 tshgchhyhnygnkvpydalifseellmkdpnyqldkadvnevkgyllkvdkyyvylkd 120
> 141 AAHANVFRKEDINOKOEHYDNKENVSNVAVARSOGRTYTNDDGVFPADIIEDTGA 200
> 121 aahadvfrkdelnqkqghvdknekvsnvavarsqgryttdndgvfnpadlledtga 180
> 201 YIVPGGHYHYIPKSDLSASELAALKAHLAGKNMOPSOLESYSTASDNNTQSVANGSTSK 260
> 181 yivpgghyhyipkdselsaseelaakahlagknmpsqysystasdnntqsvangstsk 240
> 241 pankeenlqslldkelydpsaqryesdglvfdpakllartpnyvalipghdhyhfiyysk 300
> 321 LSALTEKIRAMVPISGTGSTVSTNAKPNEVVSLSGSLSPSITTSKELSSASDGYIFN 380
> 301 lsaleeklarmvpisgtgstvttnakpnevvsisglspsitskelsasdgylfn 360
> 381 PRDIYEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATSPSLPINPGTSHEKHEED 440
> 361 pkdiveetataiyvrhgdhfyhlypknsqigqplpnnsiatpsslpinpgtshshekheed 420
> 441 GTFEDANRIIADESGFVNSHGDHNYFFKDKLTREQIQAQAQHLIEVYTSNIGLDSLS 500
> 421 gtfednriiaedeegfvnshgdhnyffkdklteeqiqaqahlievtsnigldslys 480
> 501 HBDDYGNKKNEMKDLKTEEKIAGIMKQYGRRESIYVNNKKNALITYHGHNNADPID 560
> 481 hbddygnknemkdlkteekiagimkqygrresiynnkknalityhghnnadpid 540
> 561 EHRPVGIGSHSYTELFEKKEGVAKKEKNVYTGELITVNVMLKKNSTNNONFTLANQ 620
> 541 ehprvgigshsytelfekkegvakkeknvytgelitvnmvmlkknstnnonftlanq 600
> 621 KRYSFEPPELEKKIGINLVKILITPDGVLYEKVSGKVFGEVGNANFELDOPILPGT 680
> 601 krysfeppelekkiginlvkilitpdgvlyekvsgkvfgevgnanfelldopylpgt 660
> 681 FRYTASKRYPEVSYDGTPTPSLAUKKASOTIFYPHAGPTYLAVNPOFVPGSTDL 740
> 661 frytasrkrypevsydgtpptpslaykkaasotifypahagptylavnpofvpgstdl 720
> 741 VRYFDEFHGNALYENNYKVGELIKLPIPKINOGTTRPAGNKIPVTFEAMANYLDNOSTYIYE 800
> 721 vryfdefhgnalyennykvgeiklpiipkinogttrpagnkipvtfamanyldnostyive 780
> 801 VPILEKENDTKPSIIPORRNKAQENSKLDEKVEPKTSEVYEKKLSSETGNSNSNTL 860
> 781 vpilekendtkpsiiiporrnkaqenskldekveepktsevyekklsetgnstnsntl 840
> 861 EEPYTPDPOEAVAKFAEYKMLENVLFNMDGTIELYPSGEVIKKNADTGEAPDN 920
> 841 eepytppoeavakfaeykmlenvlfnmogtieleypsevikknadtgeapdn 900
> 921 GENKSENGKRVSTGVENOPTENKRPADSLPEAPNEKRVPEKSTONGMLNPGNVSQSDM 980
> 901 genksengkrvstgvtenoptenkrpadslpeapnekrvpekstongmlnpgnvsqsdm 960
> 981 LDPALAEAPVDPVOEKEKFTASIGLDSYTFNMKGTEIELRPSGEVIKKNLSDPIA 1039
> 961 ldpalaeapvdpvoekektasigldsytfnmkgteielrpsgevikknlsdopia 1019

```

RESULT 3
 AAB12751
 ID AAB12751 standard; Protein; 1019 AA.
 AC
 XX AAB12751;

```

> 21-NOV-2000 (first entry)
> DE Streptococcus pneumoniae strain SP64 BVH-3 protein antigen.
> XX
> DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
> KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
> KW otitis media; pneumonia; immunisation; bactericidal.
> XX
> OS Streptococcus pneumoniae.
> XX
> PN M0200039299-A2.
> PD
> XX 06-JUL-2000.
> XX 20-DEC-1999; 99WO-CA01218.
> XX 23-DEC-1998; 98US-0113800.
> XX
> PA (BIOC-) BIOCHEM PHARMA INC.
> PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
> DR WPI; 2000-453397/39.
> XX
> PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
> PT otitis media, bacteraemia and/or pneumonia
> XX
> PS Disclosure; Flg 11; 106pp; English.
> XX
> CC The present invention describes nucleic acids (I) encoding protein
> CC antigens (II) from Streptococcus pneumoniae. The protein antigens
> CC have bactericidal activity. The nucleic acids, encoding the protein
> CC antigens, may be used for the recombinant production of the proteins
> CC they encode. The protein antigens may then be used as vaccines for the
> CC prevention and treatment of Streptococcal infections in mammals
> CC (especially humans) which result in, e.g. meningitis, otitis media,
> CC bacteraemia and/or pneumonia. The present sequence represents a
> CC S. pneumoniae BVH-3 protein antigen, from the present invention.
> XX
> SQ Sequence 1019 AA;

```

Query Match 98.3%; Score 5315; DB 21; Length 1019;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

> QY 21 CAYALNHRSGENKKNRVSYYDSQSOSKSENLPDOVSQREGIOAEQIVYIKITDQGY 80
> 1 cayalnhqrsgeknknrvsyvdsqsgsksealtpdqvsqkqegiaeqivlkltdggy 60
> 81 TSHGCHYHYNKNKVPYDALFSEELMKDPNYOLKADLYNEKGGYIIKVGCKYVYVYKLD 140
> 61 tshgchhyhnygnkvpydalifseellmkdpnyqldkadvnevkgyllkvdkyyvylkd 120
> 141 AAHANVFRKEDINOKOEHYDNKENVSNVAVARSOGRTYTNDDGVFPADIIEDTGA 200
> 121 aahadvfrkdelnqkqghvdknekvsnvavarsqgryttdndgvfnpadlledtga 180
> 201 YIVPGGHYHYIPKSDLSASELAALKAHLAGKNMOPSOLESYSTASDNNTQSVANGSTSK 260
> 181 yivpgghyhyipkdselsaseelaakahlagknmpsqysystasdnntqsvangstsk 240
> 241 pankeenlqslldkelydpsaqryesdglvfdpakllartpnyvalipghdhyhfiyysk 300
> 321 LSALTEKIRAMVPISGTGSTVSTNAKPNEVVSLSGSLSPSITTSKELSSASDGYIFN 380
> 301 lsaleeklarmvpisgtgstvttnakpnevvsisglspsitskelsasdgylfn 360
> 381 PRDIYEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATSPSLPINPGTSHEKHEED 440
> 361 pkdiveetataiyvrhgdhfyhlypknsqigqplpnnsiatpsslpinpgtshshekheed 420

```

441 GIGFDANRIIAEDSGFVMSHGDHNNHYFFKKDLTEEOIKAAQKHLEEVTSNGLDLSLS 500
 421 gvgfdanriiaedsgfvmshgdnhyffkkdlteeqikaaqkhleevtsnngldsls 480
 501 HEODYDGNKEMKDLDKTIEEKTAGIMKGYKRESVYNNKRNATIIYHGHHAHPD 560
 481 heodydgnkemkdlkkleektagimkgykreseiynvkenalliyphgdhhaapd 540
 561 EHKPVGIGHSHSNYELFKPEEGVAKKEGNKYTGEEELTVNNVLKKNSTNNQNTLANCO 620
 541 ehkpvigishsnhyelfkpegvakkegnkytgeeltvnnvlkknstnnqntlanq 600
 621 KRVSEFPPELEKKIGIMLVKLTIPDGKYLEKVSQVGEVGNANFELDQPYLPQGT 680
 601 krvsfeppelekkigimlvkltipdgkylevsgkvgevgnanaelfdqpylpqgt 660
 681 EFKTIASKDYPEVSYDGTFTVPSLAKMASQTIFFPHAGDTYLRVNPQFAPKGTDL 740
 661 fkytiaskdypevsydgftftvpslaymasqtlfypfhagdtlyrvnpqfapkgtdal 720
 741 VRVDFEFGHNAVLENNYKGEIKLPIPKINOGTTRAGNKIPTFMANAYLDNOSTYIYE 800
 721 vrvdfefghnavleennykgeiklpiPKINOGTTRAGNKIPTFMANAYLDNOSTYIYE 780
 801 VPILEENOTDKPSIIPQFRKNAQENSKLDERVEEPTSEKVEKEKLESTGNSTNSNTL 860
 781 vpileenotdkpsiiPQFRKNAQENSKLDERVEEPTSEKVEKEKLESTGNSTNSNTL 840
 861 EEPYTPDPOEVAFAEYAGKLENNVLENNMGITIELYIPSEVYIKKNADTGEAPQGN 920
 841 eepytPDPoeVAFaeYAGKLENNVLENNMGITIELYIPSEVYIKKNADTGEAPQGN 900
 921 GENKSENGKVTGTVENQTPENKPADSLPEAPNEKPYVEKSTONGMLNPGNNGSDPM 980
 901 genksengkvtgTVENQTPENKPADSLPEAPNEKPYVEKSTONGMLNPGNNGSDPM 960
 981 LDPALAEAPVADPOEKEKLEKFTASYGLGDSYIFNMKGITIELRLPSGEVYIKKNLSDPTA 1039
 961 ldpalaeapvADPOEKEKLEKFTASYGLGDSYIFNMKGITIELRLPSGEVYIKKNLSDPTA 1019

RESULT 4

AAB12750 standard: Protein; 1019 AA.
 AAB12750;

21-NOV-2000 (first entry)

Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
 otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

MO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA01218

23-DEC-1998; 98US-0113800

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis.

PT otitis media, bacteremia and/or pneumonia
 XX Disclousure; Fig 11; 106pp; English.
 PS
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteremia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 CC
 XX Sequence 1019 AA;
 SQ
 Query Match 98.0%; Score 5297; DB 21; Length 1019;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1016; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 21 CAYALNORSOENKNNRVSYVDGSSQSKSENTLPDOVSOKEGTOAQIYIKITDGYV 80
 1 cayaInqhrsgenKdnrvsyvdgssqsksenlPdqvsqkegIqaeqIvklIdqgyv 60
 81 TSHGDHYHYNGKRYVDALFSEELMKDPTOLKADIVNEVKGYIIRKVGKYYVYLD 140
 61 tshghyhyngkryvDAlfseellmkDpntolkADivnevkgyiIRKvgkyyvYld 120
 141 AAHADNVRTKDEINQKQEHVAKNEKVNNAVARSGRRTYTDGTVFNPADITIEDTGA 200
 121 aahadnvtKdeInqKqehvAKneKvNNAVARSGRRTYTDGTVFNPAdITIEDtga 180
 201 YIVPBGHYHYTPKSDLSASELAAKAKLACKNNQPSOLASVSTADNNQTVAKGTSK 260
 181 yivpbgHyhytpKsdlsASElAAKakALacknnQpsolASVstAdnnQTVakGtsk 240
 261 PANKSENOQLKELDYDSPSAQRSESDGLVFDPAKIIISRPNGVAIYHGHYHPIYSK 320
 241 panKsenoqlKELdyDspSAQRSESDglVfdPAkIIISrPngvAIYhghYhPIysk 300
 321 LSALEKTAARVPIISGTSVSTNAKPNVSVSSGSSSNPSLTSKELSSASDGYTFN 380
 301 lsaLeekTAArVPIISGtsVstNAKpNVsvSSGssSNpslTSKELssASdgyTFn 360
 381 PKDIVEEATAYIYRHGDHRYIKRSNOIGOPTLPNSLAPPSLPINPTSEKHEED 440
 361 pKdivEEatayIYrhgdhryIKrsNOIGOPTLPnSLAPPSLPINPTSEkHEED 420
 441 GYGF DANRIIAEDSGFVMSHGDHNNHYFFKKDLTEEOIKAAQKHLEEVTSNGLDLSLS 500
 421 gvgfdanriiaedsgfvmshgdnhyffkkdlteeqikaaqkhleevtsnngldsls 480
 501 HEODYDGNKEMKDLDKTIEEKTAGIMKGYKRESVYNNKRNATIIYHGHHAHPD 560
 481 heodydgnkemkdlkkleektagimkgykreseiynvkenalliyphgdhhaapd 540
 561 EHKPVGIGHSHSNYELFKPEEGVAKKEGNKYTGEEELTVNNVLKKNSTNNQNTLANCO 620
 541 ehkpvigishsnhyelfkpegvakkegnkytgeeltvnnvlkknstnnqntlanq 600
 621 KRVSEFPPELEKKIGIMLVKLTIPDGKYLEKVSQVGEVGNANFELDQPYLPQGT 680
 601 krvsfeppelekkigimlvkltipdgkylevsgkvgevgnanaelfdqpylpqgt 660
 681 EFKTIASKDYPEVSYDGTFTVPSLAKMASQTIFFPHAGDTYLRVNPQFAPKGTDL 740
 661 fkytiaskdypevsydgftftvpslaymasqtlfypfhagdtlyrvnpqfapkgtdal 720
 741 VRVDFEFGHNAVLENNYKGEIKLPIPKINOGTTRAGNKIPTFMANAYLDNOSTYIYE 800
 721 vrvdfefghnavleennykgeiklpiPKINOGTTRAGNKIPTFMANAYLDNOSTYIYE 780

2Y VPILEKENDTKPSIIIPQFRNKAQENSKLDEKVEEPTSEKVEKEKLESETGNSNSTL 860
 781 VPILEKENDTKPSIIIPQFRNKAQENSKLDEKVEEPTSEKVEKEKLESETGNSNSTL 860
 2Y 861 EEPYTPDVOEKAKFAESYGMKLENTLFNMDGTIELYIPSGEVYIKKMAFTGEAPQGN 920
 841 eepytvdvqekvakaesymklienlfnmdgtielypsgevyikkmaftgeapqgn 900
 2Y 921 GEKPPSENGKVSIGTYENQPTENKRPADSLPEAPNEKVPKRENTDNGKLNPEGNVSDPM 980
 901 genkpsengkvsigtyenqptenkpadslpeapnekvprentdngklnpegnvsdpm 960
 2Y 981 LDPALAEAPAVDVOEKLEKFTASYGGLDSYVFNNMGTELRPLPGEVYIKKMLSDPIA 1039
 961 ldpaleeapavdvqeklekftasygldsyvfnnmgtelelrpsevyikkmlsdpia 1019
 RESULT 5
 AAB12748
 ID AAB12748 standard; Protein; 1019 AA.
 AC AAB12748;
 DT 21-NOV-2000 (first entry)
 DE Streptococcus pneumoniae strain W02 BVH-3 protein antigen.
 KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KN prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 OS Streptococcus pneumoniae.
 PD WO200039299-A2.
 06-JUL-2000.
 20-DEC-1999; 99MC-CA01218.
 23-DEC-1998; 98GS-0113800.
 (BIOC-) BIOCHEM PHARMA INC.
 PT Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 WPI; 2000-452397/39.
 Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteraemia and/or pneumonia -
 Disclosure; Fig 11; 106pp; English.
 The present invention describes nucleic acids (i) encoding protein
 antigens (ii) from streptococcus pneumoniae. The protein antigens
 have bactericidal activity. The nucleic acids, encoding the protein
 antigens, may be used for the recombinant production of the proteins
 they encode. The protein antigens may then be used as vaccines for the
 prevention and treatment of Streptococcal infections in mammals
 (especially humans) which result in, e.g. meningitis, otitis media,
 bacteraemia and/or pneumonia. The present sequence represents a
 S. pneumoniae BVH-3 protein antigen, from the present invention.
 Sequence 1019 AA;
 SQ

Query Match 97.9%; Score 5291; DB 21; Length 1019;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

21 CAYALNHRQSEKNNRNVSYVDSOSSQSEKSENLTPDOVSOKEGIAEOIVIKITDGYV 80
 1 cayalnhrsqseknrvsyvdsossqseksentlpdovsokegiaeovikitdgyv 60

81 TSHGDHYHYNGKVPYDALFSEELMKDNPYOLKADIVNEVAGGIIIVDGKYYVYLED 140
 61 tshgdhyhyngkvydal fseellmkdnpylkdadivnevaggyiivdgkyvyyled 120
 141 AARADNVRKDELINRKOQEHVNDKENVNSNVAVARSGYRTTNDGVFEPADIEETGNA 200
 121 aaradnvrkdelnrkqehvndkenvnsnvaavarsgyrttndgvfepadieetgna 180
 201 YIYPHGCHYHYIPKSDLSASELAARAKHLAGKMKQPSQSLSYSTADNNTQSAKSGTAK 260
 181 yiyphgchyhyipksdlsase laarakhlagkkmqpsqslsystadnntqsa ksgtak 240
 261 PANKSENLQSLKELDYSPSAQRYSESDLVDPAKIISRTPNGVALPHGDHYHFIPIYK 320
 241 pankse nlqslkelydsp saqryse sdlvdpakiisrtpngvalphgdhyhfiipyk 300
 321 LSALEEKIARMPISGTGSTVSTNAKPNEVYSSLSGSSNPSSTLSKELSSASDQYIN 380
 301 lsaleeki armpisgtgstvstnakpnevys slsgssnpsstlskelssasdqy in 360
 381 PHDIVETATAYIVRHGDHPHYIPKSNQIGOPTLPNNSLATPSPSLPIPGTSHKHEED 440
 361 phdivetatayivr hgdhphyipksnqigoptlpnnslatpssplpigtshkhe ed 420
 441 GYGF DANRITIAEDSEGFVASHGDHNYFFKDKLTREQIKAAOKHLEEVYTSNGLDSSL 500
 421 gygf danritiaedse gfvashgdhnyffkdklteeqikaokhleevytsn gl dssl 480
 501 HEODYRGNKEMKMDLKIETKTAGIMKOYKVRRESIVYNNKKNALITPHGDHNDPID 560
 481 heodyr gnkemkmdlki etktagimko ykvrresiv ynnkknalitphgdhndpid 540
 561 EHKPVGIGSHSNVTELFKPEGVAKKRGKNVYTGSELTNVNLLKSTENNONFTLANGO 620
 541 ehkpvigishsnvte lfkpegvakrgknvytgse ltnvnllkstennonftlang o 600
 621 KRVSFSPPELEKRIKGINMLVKTITPDGKYLEKYSKVEEGVGNLANEELDOPYLPQGT 680
 601 krvsfsppelekr ikginmlvktitpdgkylekyskveegvgnlan eeldopylpqgt 660
 681 FKYTIAKDYPEVSYDGTFTVPTSLAYKMAQSOTIYFPHAGDTYLRVNPQFAVFKGTDAL 740
 661 fkytia kdypevsy dgtftvptslaykmaqsoti yfphagdtylr vnpqfavfkgt dal 720
 741 VRVDFEHRGNAYLENNYKVEIKLPIPKLNGCTTRRGNKIPPTFMANAYLDNQSYIYE 800
 721 vr vdfehr gnayle nnykveiklpi pklngcttrr gnkipptfmanayldnqsyiye 780
 801 VPILEKENDTKPSIIIPQFRNKAQENSKLDEKVEEPTSEKVEKEKLESETGNSNSTL 860
 781 vpilek endtkpsii ipqfrnkaqen skldekvee ptksekekekle setgnsnstl 840
 2Y 861 EEPYTPDVOEKAKFAESYGMKLENTLFNMDGTIELYIPSGEVYIKKMAFTGEAPQGN 920
 841 eepytvdvqekvakaesymklienlfnmdgtielypsgevyikkmaftgeapqgn 900
 2Y 921 GEKPPSENGKVSIGTYENQPTENKRPADSLPEAPNEKVPKRENTDNGKLNPEGNVSDPM 980
 901 genkpsengkvsigtyenqptenkpadslpeapnekvprentdngklnpegnvsdpm 960
 2Y 981 LDPALAEAPAVDVOEKLEKFTASYGGLDSYVFNNMGTELRPLPGEVYIKKMLSDPIA 1039
 961 ldpaleeapavdvqeklekftasygldsyvfnnmgtelelrpsevyikkmlsdpia 1019

Query Match 97.9%; Score 5291; DB 21; Length 1019;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

21-NOV-2000 (first entry)

Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.
 Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 otitis media; pneumonia; immunisation; bactericidal.
 Streptococcus pneumoniae.
 WO200039299-A2.
 06-JUL-2000.
 20-DEC-1999; 99WO-CA01218.
 23-DEC-1998; 98US-0113800.
 (BIOC-) BIOCHEM PHARMA INC.
 Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 WPI; 2000-452397/39.
 Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteraemia and/or pneumonia -
 Disclosure; Fig 11; 106pp; English.
 The present invention describes nucleic acids (I) encoding protein
 antigens (II) from Streptococcus pneumoniae. The protein antigens
 have bactericidal activity. The nucleic acids, encoding the protein
 antigens, may be used for the recombinant production of the proteins
 they encode. The protein antigens may then be used as vaccines for the
 prevention and treatment of Streptococcal infections in mammals
 (especially humans) which result in, e.g. meningitis, otitis media,
 bacteraemia and/or pneumonia. The present sequence represents a
 S. pneumoniae BVH-3 protein antigen, from the present invention.
 Sequence 1019 AA:

Query Match 97.9%; Score 5291; DB 21; Length 1019;

Best Local Similarity 99.5%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;

Matches 1014; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

21 CAYALNQRSEKQKNNRVSYYDSQSSQKSENLTPDVOSEKIGIAQIYIKTDQGYV 80
 1 cayalnqrseknknrvsvydsqssqsksenltpdvsgsqegiaeqiyiktdqgyv 60
 81 TSHGDHYHYNGKVPYDALFSEBLMKDPNTQKADIVNEVKGYYIIRVDGKYYVYIKD 140
 61 tshgdhyhyngkvydalfselmlkdpntqkadvnevkgyyilrvdgkyyvyikd 120
 141 AAHADNVTKQKINNOKEHYKDNKENVSNVAVASQGRYTNQGYVNPADIIEDTQNA 200
 121 aahadnvtkkeinnokehkdnkenvsnvavasqgrytndgyvnpadiliedtqna 180
 201 YIVPHGHHYTPKSDLSASELAAKAHIAKRNQPSQLSYSTASDNNQTSVAKGSGSK 260
 181 yivphghhytpksdlsaseelaakahiakrnqpsqlsystasdnntsvakgsgsk 240
 261 PANKSENTOSLKLKLYDSFAQRYSESDGLVDPAAKISRTPNQVALPHGHHYTPISK 320
 241 pankseentoslklklydsfraqryseedglvdpakisrtpnqvalphghhytphysk 300
 321 LSALEEKIARAVPISGSGSTVSTNAKPREVVSISGSSNSSTKESLSASDGIYFN 380
 301 lsaleekiaravpissgsgstvtanakprevvsssgssnsstkeslsasdgifyfn 360
 381 PKDIYEETATAYIVYHGHFHYIPKSNQIGOPTIPNNSLAPPSPLPNTPTSEKHEED 440
 361 pkdiyeetatyivyhghfhyipksnqigoptipnnslappsplpntptsekhedd 420
 441 GYGFDAANRIAEDESGFYMSGDHNYFFKDLTEQIKAAQKHLLEVYKTHNGIDSLSS 500

|||||
 421 gygfdaanriaeedsgfmsghnhyffkdlteeqikaaqkhlveekshngidslss 480
 501 HEOPYGNAKMKDLDRKIEKIGIMKQGVKRESTVYNNKXNAITYPHGDHHAADPID 560
 481 heopygnakmkdldrkiekigimkqvkrrestvynnkxnaityphgdhhaadpid 540
 561 EHKPVGISHSHNTELEFEREGVAKKEGNKYTTEGELTNVNLKNSFTNNQNTFLANGQ 620
 541 ehkpvgishshnteleferegvakkegnkyttegelnvnlknsftnnqntflangq 600
 621 KRVSFPEPELEKKIGIMLVKLTTPDGKYLEKVSQKVGSGVGNANFELDQYLPQGT 680
 601 krvsfpepelekkigimlvklttpdgkylevsgkvsgvgvgnanfelldqyilpqgt 660
 681 FKYTASADYPEVSGDGFPTSLAYKASOTIFYPHADDTLRVROPAYKGTDAL 740
 661 fkytasadypevsdgdfptslaykasotifyphaddtlrropaykgtdal 720
 741 VRVDFEFGNAYLENNKYVGEIKLPIPLNQTTRTAGNKIPYFMANAYLDNSTYIVE 800
 721 vrdfefgnaylennkyvgelklpplngttrtagnkipyfmanayldngstyive 780
 801 VPILEKENQTPKPSILPOFKRNKAQENSKLDEKVEBPRTSKVEKEKISFTGNTSNTL 860
 781 vpilekenqtkpsilpqfkrnkqenskldkveepkrtsekekeksftgntsnstl 840
 861 EEVPTVPDVOEKVAFKFAESYGMKLENYLFENDGTIELYPSGEYIKKMADEFTGAPQGN 920
 841 eevptvpdvokvafkfaesygmkleynyfendgtielylpsgeyikkmadefteapqgn 900
 921 GENRSENGKYSTGTVENQPTENRPAISLPEAPPEKPYKPNSTDNGLNDEGNVSDPM 980
 901 genrsengkystgtvenqptenrpaislpeapekpykpnstdnnglndegnvsdpm 960
 981 LDPALKEAPANDYQOEKIEKFTASYGIGLSDVITRMDGTIELRPSGVIYKKNISDPIA 1039
 961 ldpaleeapandvyokeiekfktasygiglsdvitrmdgtielerpsgviyknisdpi 1019

RESULT 7

AAB12752
 ID AAB12752 standard; Protein: 1019 AA.

AC AAB12752;

DE 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae strain P421 BVH-3 protein antigen.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

XX otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

PN WO200039299-A2.

PD 06-JUL-2000.

XX 20-DEC-1999; 99WO-CA01218.

XX 23-DEC-1998; 98US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.

XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

XX WPI; 2000-452397/39.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX

15 Disclosure: Fig 11; 106bp; English.

16 The present invention describes nucleic acids (I) encoding protein
17 antigens (II) from Streptococcus pneumoniae. The protein antigens
18 have bactericidal activity. The nucleic acids, encoding the protein
19 antigens, may be used for the recombinant production of the proteins
20 they encode. The protein antigens may then be used as vaccines for the
21 prevention and treatment of Streptococcal infections in mammals
22 (especially humans) which result in, e.g. meningitis, otitis media,
23 bacteraemia and/or pneumonia. The present sequence represents a
24 S. pneumoniae BVH-3 protein antigen, from the present invention.

25 Sequence 1019 AA;

Query Match 97.9%; Score 5291; DB 21; Length 1019;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 21 CAYALNHRSGENKNNRVSYSOSOSKSENTPPOVOSKEGIAQOIVYKIRIDOGYV 80
2 1 cayalnhtrsgenknnrvsysydsqsgskentppdqvsqegivkiktdggvv 60
3 TSHGDHYHYNGKVPYDALFSEELMKDPNTOLKADIVNEVKGYYIRKVDGKYYVYVLD 140
4 61 tshgdhyhyngkvpdydalffseellmkdpnyqlkdadlvnevkgyllkvdgkyyvylkd 120
5 141 AAHANVTKDEINQOKDHYKDKNKVSNVAVASOGRTTNDGYVNPADITDGTGNA 200
6 121 aahadvtkdeinrqkghvkdnekvsnvavarsgrytndgyvnpaditdgtgna 180
7 201 YVPRGHGHHYIPKSDLSASELAAKAAHLAGKNNKOPOLSSTSTSDNNKTSVAAGSTSK 260
8 181 ylvprghghyhyipkdsdlsaseelaakahlagknnkopsqstsdannktsvagsk 240
9 261 PANKSENTLOSLKELYDPSAORSSESDLVDPAAKISRTPNGALPHGHHYIPYSK 320
10 241 pankeentloslkelydpsaorsesdglvdpaklsrtpngvalphghhytlypsk 300
11 321 LSALEEKIARVPISGTSYSTNAKPNHYVSSLSGSLSSNPSLTTSELSASASGTYFN 380
12 301 lsaleekiarvpisgtsystnakpnhvsslsagslssnpsltskelsasasgtyfn 360
13 381 PRDIYEETATAYIYHSGHGFHYIPKSNIOGPTLPNNSLAPSPSLPFPNGSHKHEED 440
14 361 pkdlyeetatyiyhsgdhfhyipksnigpctlpnnslapspslpfpngshkheed 420
15 441 GYGDPANRIILEDSESGFVMSHGDHNFHFKDLTEEOIKAOKHLEEKYKSHNGDLSLS 500
16 421 gygdpanriiledesegfvmshgdhnhffkdlteeqkaeqkhlseektshngdls 480
17 501 HEODYPGAKEMKDLKIEKIAKIMKOYVAKRESYVYNEKNAIYYPHGDHHDHAPID 560
18 481 heodypgakemkdlkiekiakimkoymvaksresyvyneknaiyypghdhdhapid 540
19 561 EHKPRGIGHSHSNELFEPREGVAKKKGKNTYTGEBELTNVNLKNSFNNOONTLNGO 620
20 541 ehkprgighshsnelfepregvakkekngkytgeeltnvnlknsfnnoontlng 600
21 621 KRVSFSPPELEKIKGIMLAKLITPDGKVEKYSKGKFGVGNIANFELDOPILPQOT 680
22 601 krvsfsfpelkekikgimlakltpdgkvekyksgkfgvgvgnianfelpilpqt 660
23 681 EKYTIASADYEVSTDGTFYPTSLAYKMASQITTFPHADDTILYNPQAPVPRKGTAL 740
24 661 kytiasadyevstdgtyfptslaykmasqitfphadgtylrvnpqavprkgtal 720
25 741 VRVDEFGNAYLNNYVGGIKLPIPLNGGRTAGNKIPVPMANAYVNDNGSTYVE 800
26 721 vrvdefgnaylennyvggiklpiplnggtrtagnkprvpmamayndngstye 780
27 801 VPILEKENGQTKPSLIPFKRNKAQENSKLDEKVEEPTSEKVEKLSGNGSTNLT 860

DB 781 vpilekengtkpsllpfkrnkagenskidekveeptsekevекlssetgnsntl 840
QY 861 EEPVDPVOEKVAKKEAESYGMKLENTLFNMDGTELYLPGEVYKKNMAFTEGAPGN 920
DB 841 eevpdpvovkavakkeesygmkleentlfnmddgtelylpgevlykknmadftega 900
QY 921 GENPSENGKYSTGVENQPTENRPAUSLPEAPNEKPYKPNSTDNKMLNPGNVGSDPM 980
DB 901 genpsengkystgvtenqptenrpadlpeapnekpykpnstndnmlnpgnvgsdpm 960
QY 981 LDPALEAPVADPVOEKLEKFTASTGGLDSVIFNMGGTILRLPGEVYKKNLSDPIA 1039
DB 961 ldpaaleapvadvovklekftastgglsvifnmddgtelrlpgevlykknlsdila 1019

RESULT 8
AAB12753
ID AAB12753 standard; Protein; 1019 AA.
XX
AC AAB12753;
XX
DF 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae; strain A66 BVH-3 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Plneau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PF Streptococcal antigens useful for vaccinating against e.g. meningitis,
PF otitis media, bacteraemia and/or pneumonia -
XX
PS Disclosure: Fig 11; 106bp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

Query Match 97.9%; Score 5291; DB 21; Length 1019;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 21 CAYALNHRSGENKNNRVSYSOSOSKSENTPPOVOSKEGIAQOIVYKIRIDOGYV 80
DB 1 cayalnhtrsgenknnrvsysydsqsgskentppdqvsqegivkiktdggvv 60
QY 81 TSHGDHYHYNGKVPYDALFSEELMKDPNTOLKADIVNEVKGYYIRKVDGKYYVYVLD 140

```

b 61 tshgdhyhnyngkvydalfseellmkdpayqldadivevkggylikvdgkyvyyikd 120
Y 141 AAADNVRTDEINROKOEHVKNDEKYNNAVAARSGRTTNDGYFNPADIIEDTGN 200
b 121 aahadavrttdelnrqgeivkdnkynsnvavaregrytlndgyfnpadliedtgna 180
Y 201 YIYPHGHHYIIPKSDISASELAAKAHLAKKNQPSQLSYSTASDNNTQSAKGSTK 260
b 181 ylyphrghnylypsdlsaseelaakahlagknmpsqqlsysslaednltqsvkgsstk 240
Y 261 PANKSEMSLLEKLYSPSAORSESDGLVDPAKIISTPNCVALPHGDHVFITYSK 320
b 241 panlsenlgellkelyspasqyseeqlytdpakiistpncvalphgdhvfitysk 300
Y 321 LSALEEKIARVPISGTVSTNAKPNEVYVSSLSGSSNPSSLTTSKELSASDGYIFN 380
b 301 lsaleekiamvpiisgsvstnakpnevvaslgsasnpsslttskelsasadyifn 360
Y 381 PKDIVEETATAYIVRHGDHNYIPKSNQIQPTLPNNSLATPSPSLPINTGSHKEHEE 440
b 361 pkdiveetacayivrhgdhnyipksnqisqptlpnnslatpsspplintgshkheed 420
Y 441 GYGDDARRITAEDESGVMSGDHNYFFKKDLTEBOIKAAOKLEEVKTSNHCGLSLSS 500
b 421 gygddarritlaedesgvmshgdhnyffkkdlteboikaaokleevktsnhs 480
Y 501 HEODPNAKEMKMDLKKIEKINGIKOYGVKNESIVVKNKNAITYPHGDHHAAPID 560
b 481 heodpnaekmdldkkiekingikoygvknesivvknknaityphgdhhaapid 540
Y 561 EHKRVGIGHSHSNELKRPBGVAKKRGKNVYTGSEELTNVNLKSTPNQNTLANGC 620
b 541 ehkrvgighshsnelkrpbgvakrgknvytgseeltnvnllkstpnngntlanng 600
Y 621 KRVSFSPPELEKELGIMLYKLTTPDGKYLEKSGVFGVGNINLEFDQYLLGQOT 680
b 601 krvsfsppelekelligimlyklttpdgkyleksgvfgvgninlefdqyllgqot 660
Y 681 FKXITASKDPREVSIDGTFVPSLAYKMASQITFYFHAGDYILRVNPPFAVKGTDAL 740
b 661 fkytiaskdprevsydgtfvpslaykmasqitfyfhagdyilrvnppfavykgtdal 720
Y 741 VRVDEHGNALYLENNKTVGEIKPIPKLNGTTRTGNKIPVTFMANAYLDNSTIIVE 800
b 721 vrvdehgnalylenntvgeikpikplngttrtagnipvtfmanayldnsgstive 780
Y 801 VPILEKNOGDKPSILFOFRKNKQENSKLDEKVEEPTSKVEKELSEFNGSTNSTL 860
b 781 vpileknogdkpsilfofrknkqenskldkveeptskvekekelsfngstnstl 840
Y 861 EEVPTVPVOEKVAKFAESYGMKLENTLFNMDGIELYLPSEGVYIKKNMADFGCEAOGN 920
b 841 eeptvpvoekvakfaesygmkleentlfnmngtlielylpsgevikknmadfgceapgn 900
Y 921 GENRPSNGKSVSTVNOPTENKPADSLPEAPNEKVKYKENSNDKNMLNPEGVSGDPM 980
b 901 genrpsngksvstvgntenpadsleapnekvkykensndknmlnpegvsgdpm 960
Y 981 LDPALBAPAVDPVOEKLEKFTASYGGLDSVTFNMMDGTIELRLPSEGVYIKKNLSDTIA 1039
b 961 ldpaleapavdpvoeklekftasysgglsvtfnmngtlielrlpsgevikknlsdtia 1019

```

```

ESTUT 9
D AB12721
X AAB12721 standard; Protein: 840 AA
C AAB12721
X X
T 21-NOV-2000 (first entry)
X Streptococcus pneumoniae SP63 BVH-3 protein antigen SEQ ID NO:16.
X

```

```

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW Propylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW Otitis media; pneumonia; immunisation; bactericidal.
OS Streptococcus pneumoniae.
PN MO200039299-A2.
PD 06-JUL-2000.
PF 20-DEC-1999; 99MO-CA01218.
PR 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
DR N-PSDB; AAA65738.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
PS Claim 16; Fig 19; 106pp; English.
XX
CC The present invention describes nucleic acids (i) encoding protein
CC antigens (ii) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae SP63 BVH-3 protein antigen.
XX
SQ Sequence 840 AA:

```

```

Query Match 79.1%; Score 4276.5; DB 21; Length 840;
Best Local Similarity 82.3%; Pred. No. 4e-251;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;
QY 21 CAYALNORSEKNNKRVSYVDGSSQSKSENITPVOVQKSGIAEOIVIKITDGYV 80
DB 1 cayalngtrsegenkdnrvsyvdgsgsqsenltpvqsqkegiqeqivikltdgyv 60
QY 81 TSHGDHYYNGVVPYDALFSEELMKDPYQLDADIVEVKGYYIIKYDGXYVYIKD 140
DB 61 tshgdhyhnyngkvydalfseellmkdpayqldadivevkggylikvdgkyvyyikd 120
Y 141 AAADNVRTDEINROKOEHVKNDEKYNNAVAARSGRTTNDGYFNPADIIEDTGN 200
b 121 aahadavrttdelnrqgeivkdnkynsnvavaregrytlndgyfnpadliedtgna 180
Y 201 YIYPHGHHYIIPKSDISASELAAKAHLAKKNQPSQLSYSTASDNNTQSAKGSTK 260
b 181 ylyphrghnylypsdlsaseelaakahlagknmpsqqlsysslaednltqsvkgsstk 240
Y 261 PANKSEMSLLEKLYSPSAORSESDGLVDPAKIISTPNCVALPHGDHVFITYSK 320
b 241 panlsenlgellkelyspasqyseeqlytdpakiistpncvalphgdhvfitysk 300
Y 321 LSALEEKIARVPISGTVSTNAKPNEVYVSSLSGSSNPSSLTTSKELSASDGYIFN 380
b 301 lsaleekiamvpiisgsvstnakpnevvaslgsasnpsslttskelsasadyifn 360
Y 381 PKDIVEETATAYIVRHGDHNYIPKSNQIQPTLPNNSLATPSPSLPINTGSHKEHEE 440
b 361 pkdiveetacayivrhgdhnyipksnqisqptlpnnslatpsspplintgshkheed 420
Y 441 GYGDDARRITAEDESGVMSGDHNYFFKKDLTEBOIKAAOKLEEVKTSNHCGLSLSS 500
b 421 gygddarritlaedesgvmshgdhnyffkkdlteboikaaokleevktsnhs 480

```


244 gysfdaurliaedesgfymshgdnhyffkdlleedqkaaqkhlvevktshngidsls 303
 301 HEODYPGNAAKEKKDLKDKIEEIKIAGIMQGYKRSIYVNEKNAIITPHGDHHAADID 560
 304 heqdypanakekmdldkklleeklaeglmkygvtreslyvneknaillphgdhhaadid 363
 561 EHKPVGIGSHSNYELFPREGVAKKEGNKYTGELTNVNLKNSFNNOFNLANGO 620
 364 ehkpygigshsnhyellfpreegvakegnkyltgseeltvnlknsfnngfclangq 423
 621 KRVSFSPPELEKKIGINKLVLTIPDGKVLKESVGFGEVGNIANEELDQPLPGQT 680
 424 krvsfsppelekkigimlvltltpdgkvlkvegkvfgevgvgnianfeldqpylpqgt 483
 681 FRYTASNDYREVSYDGFYVPTSLAYKASQTIYPPHAGTYLAVNPQFVPGSTAL 740
 484 fkytlaskdyrevsydgftvptslaykmaagcllypfnagctlylvnpqfsvpgstald 543
 741 VRFVDFGNATLENNYKVEIKLPIPKLNQGTTRAGNKIPVTFMANAYLDNOSTYIVE 800
 544 vrvfdefngnylennhykvgelkplpklngtctrtagnklpvtmanayldngstyive 603
 801 VPILEKNOTDKPSILPQFRNKAQENSKLDEKVEEPTSEKEREKLSGTGNSNSTL 860
 604 vpllekngtdkpsilpqfkrnkagenaekldkveektsekveeklsctgnsanstl 663
 861 EEVPPVDPQVEKVAAPASYSYGMKLENTLFPNDGTIELYLPSEGVKKNADTGAPOGN 920
 664 eeppvdpqvekvakeevaygmkleentlfpndgtielylpsegvakknadtgaepgn 723
 921 GENKSENGKYSTGTVENOPTENKPADSLPEAPNEKPYKPEKSTNGMLNPGVNSDPM 980
 724 genksengkystgtveopktenkpadslpeapnekykpekstngmlnpgvnsdpm 783
 981 LDPALAEAPVDPVOEKELEKFTASTYGLGIDSVTFNNDGTIELRLPSGEVYKKNL 1034
 784 ldsaleapavdpvogekelektastystyglidsvtfndgtielrlpsgeviknll 837

RESULT 10
 AAB12744
 AAB12744 standard; Protein; 780 AA.

AAB12744;
 21-NOV-2000 (first entry)
 Streptococcus pneumoniae NEM15 protein antigen SEQ ID NO:78.
 Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
 otitis media; pneumonia; immunisation; bactericidal.
 Streptococcus pneumoniae.
 MO200039299-A2.
 06-JUL-2000.
 20-DEC-1999; 99WO-CA01218.
 23-DEC-1998; 98US-0113800.
 (BIOC-) BIOCHEM PHARMA INC.
 Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 WPI; 2000-452397/39.
 Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteremia and/or pneumonia
 Claim 18; Fig 43; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the proteins
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteremia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae NEM15 protein antigen.
 XX
 SQ Sequence 780 AA;
 Query Match 75.7%; Score 4090; DB 21; length 780;
 Best Local Similarity 100.0%; Pred. No. 7,2e-240;
 Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 21 CAYALNQRSEENKDNKNVSYVDSOSSOKSENTLPDOVSQKEGIQABQIVIKIDQGY 80
 1 cayalnqrsgenkdnknvsvydsqsgskenlppdqvsgkqgaqvialkltdgvy 60
 81 TSHGDHYHYNGKVPYDALFSEELAKDPNTQKADIVNEVKGCIITKDGKYYVILKD 140
 61 tshgdhyhyngkvydalfeellakdpnyqkldadivnevkggylkvdgkyvvyikd 120
 141 AAHADNVPTKDEINKOKEHYVDNEKVNVAASOGRTYTNDDGYVNPADIIEDTGNA 200
 121 aahadvptkdeinkokehyvdnekvnsnvaasogrtytnddgyvnpadiiedtgna 180
 201 YIVPGGHYHYTPKSDLSASELAARLAKGNMOP SOLSYSTASDNNQTSVAKGSTSK 260
 181 yivpgghyhytpksdlsaseelaakhlagknmpsglsyastadnntgvakgstsk 240
 261 PANKSENTQSLKEIYDSPAQRYSQGLVDPKAIISRTNGVAILPHGDHYHITPSK 320
 241 pankseentqslkeiydspaqrysglvdvpkaiisrtngvailphgdhyhitpsk 300
 321 ISALEKIRARVPISGTGSTVSTNAKPNEVVSLSLSNPSSLTTSKELSSASDGIYFN 380
 301 lsaleekirarvpisgtsvstnakpnevvsjslsenpsltskelsasasdgifyfn 360
 381 PRDIVEETATAYVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSEKHEED 440
 361 prdiveetatayivrghdhyipksngtqptlpnnslatpsslpinpgtshkheed 420
 441 GYGFDAURLIAEDES GFYMSHGDHNYFFKDLTEBQTKAAKHLIEVKTSHNGIDSLSS 500
 421 gysfdaurliaedesgfymshgdnhyffkdltebqtkaaqkhlvevktshngidsls 480
 501 HEODYPGNAAKEKKDLKDKIEEIKIAGIMQGYKRSIYVNEKNAIITPHGDHHAADID 560
 481 heodypgnakekmdldkklleeklaeglmkygvtreslyvneknaillphgdhhaadid 540
 561 EHKPVGIGSHSNYELFPREGVAKKEGNKYTGELTNVNLKNSFNNOFNLANGO 620
 541 ehkpygigshsnhyellfpreegvakegnkyltgseeltvnlknsfnngfclangq 600
 621 KRVSFSPPELEKKIGINKLVLTIPDGKVLKESVGFGEVGNIANEELDQPLPGQT 680
 601 krvsfsppelekkigimlvltltpdgkvlkvegkvfgevgvgnianfeldqpylpqgt 660
 681 FRYTASNDYREVSYDGFYVPTSLAYKASQTIYPPHAGTYLAVNPQFVPGSTAL 740
 661 fkytlaskdyrevsydgftvptslaykmaagcllypfnagctlylvnpqfsvpgstald 720
 741 VRFVDFGNATLENNYKVEIKLPIPKLNQGTTRAGNKIPVTFMANAYLDNOSTYIVE 800
 721 vrvfdefngnylennhykvgelkplpklngtctrtagnklpvtmanayldngstyive 780

RESULT 11
 AAB12731

D AAB12731 standard; Protein; 568 AA.
 X AAB12731;
 X 21-NOV-2000 (first entry)
 X Streptococcus pneumoniae NEW1 protein antigen SEQ ID NO:64.
 X Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 X prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 X otitis media; pneumonia; immunisation; bactericidal.
 X Streptococcus pneumoniae.
 X MO200039299-A2.
 X 06-JUL-2000.
 X 20-DEC-1999; 99MO-CA01218.
 X 23-DEC-1998; 98US-0113800.
 X (BIOC-) BIOCHEM PHARMA INC.
 X Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 X WPI; 2000-452397/39.
 X Streptococcal antigens useful for vaccinating against e.g. meningitis,
 X otitis media, bacteraemia and/or pneumonia -
 X Claim 18; Fig 29; 106pp; English.
 X The present invention describes nucleic acids (I) encoding protein
 X antigens (II) from Streptococcus pneumoniae. The protein antigens
 X have bactericidal activity. The nucleic acids, encoding the proteins
 X antigens, may be used for the recombinant production of the proteins
 X they encode. The protein antigens may then be used as vaccines for the
 X prevention and treatment of Streptococcal infections in mammals
 X (especially humans) which result in, e.g. meningitis, otitis media,
 X bacteraemia and/or pneumonia. The present sequence represents the
 X S. pneumoniae NEW1 protein antigen.
 X Sequence 568 AA;

Query Match 54.6%; Score 2953; DB 21; Length 568;
 Best Local Similarity 100.0%; Pred. No. 3.8e-171;
 Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

472 DLTEEOIKAAQKLEEVKTSNGSDLSSEHODYPGNAKEMKLDKRIEKGIMKQY 531
 1 dlteeqkaagkhleevktsngsdlssehodypgnakemkldkriekgimkqys 60
 532 VKRESIVYNNKNAIIPDGDHDAIDIEKPVGICHSSNTELPREGVAKKEKNV 591
 61 vkresivnnknaaiipdgdhdaidiekhpvigichssntelpregvakkeknv 120
 592 YTGELNVNVLKNSFNNOFLANGOKRVSPFPELEKKGIMLVKLTTPDQKVL 651
 121 ytgelnvnvlknsfnnoflangokrvspfpelekkkgimlvklttppdqkvl 180
 652 EKVSQVFGSGVGNIANFELDQPTLPQTFKTYTASNDYEVSTGTFYPTSLAYMAS 711
 181 ekvsqvfsgsgvgnianfelldqptlpqtfktytasndyevstgtyptslaymas 240
 712 QTFYPPHAGDTYLRNAPORAVRKGTDALYRVDEHGNALYLENNYVGEIKLPIPLNQ 771
 241 qtfypphagdtylrnaporavrkgtDALYRVDEHGNALYLENNYVGEIKLPIPLNQ 300
 772 GTTFATANKIPVFNMAVAVDNOSTYVEVPILEKENOTKPSILPQFKNKAOENSKLD 831
 301 gttfatankipvfnmaavadnostyvevpilekenotkpsilpqfknkAOENSKLD 360

QY 832 EKVEEPTSEKVEKEKLSSTNGSTNSLTLEEVPTDVOEKVAKFAASYENKLENTLFNM 891
 DB 361 ekveeptsekevkeklsstngstnsltleevptdvoekvakfaasyenkleentlfnm 420
 QY 892 DGTIELYPSGEVTKKMAADFTGEAPGNGENKPSBNGKYSTGVENQPIENKPADSLPE 951
 DB 421 dgtielypsgevtkkmaadftgeapngngenkpsbngkystgvnqpienkpadslpe 480
 QY 952 APNEKPYKPNSTDNKMLNPEGNVGSDPMDPLAEAPAVDPVOEKLEKFTASVGLSDS 1011
 DB 481 apnekpykpnstdnkmlnpegnvgsdpmplaeapavdpvoeklekftasvglids 540
 QY 1012 VIFNMDGTIELRLPSGEVTKKNSLDFIA 1039
 DB 541 vifmngdtielyrlpsgevtkknsldfia 568

RESULT 12

AAB12725
 ID AAB12725 standard; Protein; 1057 AA.

AC AAB12725;
 AC 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae NEW12 protein antigen SEQ ID NO:58.

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

PN MO200039299-A2.

PN 06-JUL-2000.

PD 20-DEC-1999; 99MO-CA01218.

PF 23-DEC-1998; 98US-0113800.

PR (BIOC-) BIOCHEM PHARMA INC.

PA Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

PI WPI; 2000-452397/39.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -

PS Claim 18; Fig 23; 106pp; English.

CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the proteins
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae NEW12 protein antigen.
 CC Sequence 1057 AA;

Query Match 54.6%; Score 2953; DB 21; Length 1057;
 Best Local Similarity 100.0%; Pred. No. 9e-171;
 Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEOIKAAQKLEEVKTSNGSDLSSEHODYPGNAKEMKLDKRIEKGIMKQY 531
 DB 1 dlteeqkaagkhleevktsngsdlssehodypgnakemkldkriekgimkqys 60

```

Y 532 VKRSIIYVNNKKNNAIIPHGHHADPIDEERPVGIGHSHSNYTELFKEPSEVAKKEGKV 591
D 61 vkreeivnnkknnaillphghhnaprdehkpvglgshshnyelfpreevakegnkv 120
Y 592 YTGELTNVNLKNSFTNNFTLANQKRVSSFPPELKKIGINLVLTIPDGKVL 651
D 121 ytgeltntvnlknstfnngnftlanqkrvssfppelekkigimlvltipdgkvl 180
Y 652 EKVSGRVGEVGNINANFELQPIIPGOTFKYTTASKDIPYVSIDGFTTVPSTAYKAS 711
D 181 ekvsgrvgevgnlanfelqpylpgotfkyltasakypvsgdfttvpstlaykmas 240
Y 712 OTIYFPFAGGTLYRVNQFAPVPGTDALVRFDEFGNATLENNYKVGELKLPRIKNO 771
D 241 otlyfpfaggtlyrvnqfapvpqtdalvrfdefgnaylennykvgelkripklnq 300
Y 772 GTRTAGNKRIFVTMANAYLDNOSTYIYEVPILEKENOTDRPSILPQKRRKKAENSKLD 831
D 301 gtrtagnkrifvtmanayldngstyivevpiilekengtdrpsilpqkrrkkaenakld 360
Y 832 EKVEEPTSEVVEKEKSESTNSNSTLEEVPTVDPVQEKYAKFAESTYGMKLENVLFNM 891
D 361 ekveeptsevekeksetnsnstleevptvdpvqekyakfaestygmklenvlfnm 420
Y 892 DGTIELVPSGEVTKNNADFTGEAPOGNGENKPSENGKYSTGVENQPTENKRPADSLPE 951
D 421 dgtielvpsgevtknnadftgeapogngenkpsengkystgvnqptenkrapdslpe 480
Y 952 APNKKPVKPNSTNGMLNPGNVGSDPMLDPALAEAPVADPVQEKLEKFTASYGLGIDS 1011
D 481 apnkkpvkpnstngmlnpgnvgsdpmldpalaeapvadvpvekletekftasyglgids 540
Y 1012 VIFNMDSIELRLSGEVIKKNLSDFTA 1039
D 541 vifmndsielrlpsgevikknlsdfta 568

```

RESULT 13

ID AAB12719 standard; Protein: 528 AA.

AAB12719;

21-NOV-2000 (first entry)

Streptococcus pneumoniae BVH-3B protein antigen SEQ ID NO:10.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia; otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

MO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99MO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N; NPI; 2000-452397/39.

N-PSDB; AAA65734.

Streptococcal antigens useful for vaccinating against e.g. meningitis, otitis media, bacteremia and/or pneumonia.

Claim 18; Fig 10; 106pp; English.

```

XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteremia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3B protein antigen.
XX
XX Sequence 528 AA;

```

Query Match 50.8%; Score 2746; DB 21; Length 528;
 Best Local Similarity 100.0%; Pred. No. 1,2e-158; Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Y 512 MKDLDRKIEEKIAGIMKOYGVKRSIYVNNKKNNAIIPHGHHADPIDEERPVGIGHSH 571
D 1 mkdlldrkieekitagimko ygvkrsi yvnnkknnaillphghhnaprdehkpvglgsh 60
Y 572 SNYELFPEEVSANKEGKVYTGSELNVNLKNSFTNNFTLANQKRVSSFPPEL 631
D 61 snyelfpreevakegnkv ytgeltntvnlknstfnngnftlanqkrvssfppe 120
Y 632 EKKIGINLVLTIPDGKVLAKVSGKVRGEVGNINANFELQPIIPGOTFKYTTASKDYP 691
D 121 ekkigimlvltipdgkvlakvsgkvrgevgnlanfelqpylpgotfkyltasakyp 180
Y 692 EVSYDGETTVPSTLAYKKAQSOTIYFPFAGGTLYRVNQFAPVPGTDALVRFDEFGNA 751
D 181 evsydgettvpstlaykkaqsotiyfpfaggtlyrvnqfapvpqtdalvrfdefgna 240
Y 752 YLENNYKVGELKLPRIKNOGTRTAGNKRIFVTMANAYLDNOSTYIYEVPILEKENOTD 811
D 241 yleennykvgelkripklnqgtrtagnkrifvtmanayldngstyivevpiilekengtd 300
Y 812 KPSILPQKRRKKAENSKLDKEVEEPTSEVVEKEKSESTNSNSTLEEVPTVDPVQOE 871
D 301 kpsilpqkrrkkaenakldkeveeptsevekeksetnsnstleevptvdpvqe 360
Y 872 KVAFKAEISYGMKLENVLFNMDSIELRLPSGEVTKNNADFTGEAPOGNGENKPSENGKV 931
D 361 kvafkaeis ygmkle nvlfmndgtielrlpsgevtknnadftgeapogngenkpsengkv 420
Y 932 STGVENQPTENKRPADSLPEAPNKKPVKPNSTNGMLNPGNVGSDPMLDPALAEAPV 991
D 421 stgvnqptenkrapdslpeapnkkpvkpnstngmlnpgnvgsdpmldpalaeapv 480
Y 992 DPVQEKLEKFTASYGLGIDSYTFNMDSIELRLSGEVIKKNLSDFTA 1039
D 481 dpvqekletekftasyglgidsytfmndsielrlpsgevikknlsdfta 528

```

RESULT 14

ID AAB12724 standard; Protein: 509 AA.

AAB12724;

21-NOV-2000 (first entry)

Streptococcus pneumoniae L-BVH-3-AD protein antigen SEQ ID NO:57.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia; otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

MO200039299-A2.

D 06-JUL-2000.
 X 20-DEC-1999; 99MO-CA01218.
 F 23-DEC-1998; 98US-0113800.
 X (BIOC-) BIOCHEM PHARMA INC.
 X Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N,
 X WPI; 2000-452397/39.
 T Streptococcal antigens useful for vaccinating against e.g. meningitis,
 T otitis media, bacteremia and/or pneumonia.
 S Claim 18; Fig 22; 106pp; English.

C The present invention describes nucleic acids (I) encoding protein
 C antigens (II) from Streptococcus pneumoniae. The protein antigens
 C have bactericidal activity. The nucleic acids, encoding the protein
 C antigens, may be used for the recombinant production of the proteins
 C they encode. The protein antigens may then be used as vaccines for the
 C prevention and treatment of Streptococcal infections in mammals
 C (especially humans) which result in, e.g. meningitis, otitis media,
 C bacteraemia and/or pneumonia. The present sequence represents the
 C S. pneumoniae L-BVH-3-AD protein antigen.

Sequence 509 AA;

Query Match 49.0%; Score 2650; DB 21; Length 509;
 Best Local Similarity 100.0%; Pred. No. 7, 6e-153;

Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKRSKYIAGSAVYISLCAYALNHRSEKDNRRVYDGSOSQSENIPTPOVS 60
 1 MKISKYLAAGSAVYISLCAYALNHRSEKDNRRVYDGSQSGSENIPTPOVS 60
 Y OKSGIOAEQIVITITDOGYTSHGDHYHNGVYPYDALPSEELMKDPYQKADIVN 120
 b 61 qkegiqaeqivlitdgytshgdhyhngvypydalseelinkdpyqkddivn 120
 Y 121 EVKGGYIIKVDGKITYYLKDAADNRKDEINRKOEHVKNEKYNVAVARSGRT 180
 b 121 evkgyyllkvdgkyyyllkdaadnrvtkdelnrqgehvkdnekynsvavarsgry 180
 Y 181 TTNDGYVFNPAIDIEETGNAYIVPBGGHYIYRKSDLSASELAAKAHLAGKMQSOLS 240
 b 181 ttndgyvfnpadliedtgnaivpbgghyhyrksdlsaseelaakahlagkmpqsols 240
 Y 241 YSSTADNNTQSVAKSSTKPKANKSENLSILKELYDSPAQRYSSESDGLVFPKATISR 300
 b 241 ysstadnntqsvaksgstkrpknksenlsgllkelydspaqryseesdglvfpkattsr 300
 Y 301 TPNGVAIPBGDHYHFIYPSKLSALEEKIARMPISGSGTVSTNAKPNEVYSSLSGLSSN 360
 b 301 tpngvailpbgdhyhfiypsklsaleekiarmpisgsgstvsnakpnevyslsglssn 360
 Y 361 PSSLTSTSKELSSSDGYIFNPKDIVEETATAYIVRHGDHNYPKNOGQPTLPNNSLA 420
 b 361 pssltstskelsssdgyifnpkdiveetataylvrhgdhnylprkngqptlpnnslla 420
 Y 421 TPSPSLPINCSTSHHEHEDGYGFDANRIIAEDSGFVSHGCHNHYFFKKDTEEQIKR 480
 b 421 tpspslpincstshhehedygfdanriiaedsgfvshgchnhyyffkkdteeqikr 480
 Y 481 AOKHLEKVTSHNGDLSLSHEDDYPGNA 509
 b 481 aqhleevtshngldslsheddygna 509

RESULT 15
 AB12723

ID AAB12723 standard; Protein; 489 AA.
 XX
 AC AAB12723;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae BVH-3AD protein antigen SEQ ID NO:56.
 XX
 DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 XX
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN MO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 XX 20-DEC-1999; 99MO-CA01218.
 XX
 XX 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N,
 XX
 PI WPI; 2000-452397/39.
 XX
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteremia and/or pneumonia.
 PT
 PS Claim 18; Fig 21; 106pp; English.

C The present invention describes nucleic acids (I) encoding protein
 C antigens (II) from Streptococcus pneumoniae. The protein antigens
 C have bactericidal activity. The nucleic acids, encoding the protein
 C antigens, may be used for the recombinant production of the proteins
 C they encode. The protein antigens may then be used as vaccines for the
 C prevention and treatment of Streptococcal infections in mammals
 C (especially humans) which result in, e.g. meningitis, otitis media,
 C bacteraemia and/or pneumonia. The present sequence represents the
 C S. pneumoniae BVH-3AD protein antigen.

Sequence 489 AA;

Query Match 47.3%; Score 2559; DB 21; Length 489;
 Best Local Similarity 100.0%; Pred. No. 2, 4e-147;
 Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 21 CAYALNHRSEKDNRRVYDGSOSQSENIPTPOVSOKSGIOAEQIVITITDOGYV 80
 b 1 cayalnhrsekenkdnrrvydvgsqsgsenltpdqvsqkegiqaeqvilitdgyv 60
 Y 81 TSHGDHYHNGVYPYDALPSEELMKDPYQKADIVNEVKGYYIIKVDGKITYYLK 140
 b 81 tshgdhyhngvypydalseelinkdpyqkddivnevkgyyllkvdgkyyyllkd 120
 Y 141 AAHADNRKDEINRKOEHVKNEKYNVAVARSGRTTNDGYVFNPAIDIEETGNA 200
 b 141 aahadnrvtkdelnrqgehvkdnekynsvavarsgrytndgyvfnpadliedtgna 180
 Y 201 YIYPBGGHYIYRKSDLSASELAAKAHLAGKMQSOLSYSSTADNNTQSVAKSSTK 260
 b 181 yivpbgghyhyrksdlsaseelaakahlagkmpqsolysstadnntqsvaksgstkr 240
 Y 261 PANKSENLSILKELYDSPAQRYSSESDGLVFPKATISRTPNGVAIPBGDHYHFIYPSK 320
 b 241 panksenlsgllkelydspaqryseesdglvfpkattsrtpngvailpbgdhyhfiypsk 300
 Y 321 LSALEEKIARMPISGSGTVSTNAKPNEVYSSLSGSSLTSTSKELSSASDGYTN 380
 b 301 lsaleekiarmpisgsgstvsnakpnevyslsglssnpsltstskelssasdgtyln 360

OY 381 PKDIVEETATAYIVRHGDHPHYIPKSNQIGOPTLPNNSLATPSPSLPINPGTSHEKHEED 440
Db 361 PKdlveetataaylvrrngdhhyipksnqigoptlpnnslatpsslpinpgtshkheed 420
OY 441 GYGF DANRIIAEDSGFVMSGHGNHYFFKKDLTEQIKAAQKHLEEVKTSHNGLSLS 500
Db 421 gygf danriiaedesgfvmshgdnhyffkkdlteeqikaqkhleevktsnhgldelss 480
OY 501 HEQDYPGNA 509
Db 481 heqdypgna 489

Arch completed: September 26, 2001, 22:30:29
Job time: 3208 sec

SEP 27 2001 09:15:57

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:30:29 : Search time 57.46 Seconds

(without alignments)
143.489 Million cell updates/sec

Title: US-09-471-255-2_COPY_15_150

Perfect score: 711
I VSLSLCAVLAALNHRSEOK.....DKRYVYLKDAAHADNVRK 136

Sequence: 1 IVSLSLCAVLAALNHRSEOK.....DKRYVYLKDAAHADNVRK 136

Search: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A.GeneSeq_0601:
1: /SIDS1/gcgdata/geneSeq/AA1980.DAT:
2: /SIDS1/gcgdata/geneSeq/AA1981.DAT:
3: /SIDS1/gcgdata/geneSeq/AA1982.DAT:
4: /SIDS1/gcgdata/geneSeq/AA1983.DAT:
5: /SIDS1/gcgdata/geneSeq/AA1984.DAT:
6: /SIDS1/gcgdata/geneSeq/AA1985.DAT:
7: /SIDS1/gcgdata/geneSeq/AA1986.DAT:
8: /SIDS1/gcgdata/geneSeq/AA1987.DAT:
9: /SIDS1/gcgdata/geneSeq/AA1988.DAT:
10: /SIDS1/gcgdata/geneSeq/AA1989.DAT:
11: /SIDS1/gcgdata/geneSeq/AA1990.DAT:
12: /SIDS1/gcgdata/geneSeq/AA1991.DAT:
13: /SIDS1/gcgdata/geneSeq/AA1992.DAT:
14: /SIDS1/gcgdata/geneSeq/AA1993.DAT:
15: /SIDS1/gcgdata/geneSeq/AA1994.DAT:
16: /SIDS1/gcgdata/geneSeq/AA1995.DAT:
17: /SIDS1/gcgdata/geneSeq/AA1996.DAT:
18: /SIDS1/gcgdata/geneSeq/AA1997.DAT:
19: /SIDS1/gcgdata/geneSeq/AA1998.DAT:
20: /SIDS1/gcgdata/geneSeq/AA1999.DAT:
21: /SIDS1/gcgdata/geneSeq/AA2000.DAT:
22: /SIDS1/gcgdata/geneSeq/AA2001.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	711	100.0	484	21	Streptococcus pneu
2	711	100.0	484	21	Streptococcus pneu
3	711	100.0	484	21	Streptococcus pneu
4	711	100.0	485	21	Streptococcus pneu
5	711	100.0	509	21	Streptococcus pneu
6	711	100.0	1039	21	Streptococcus pneu
7	687	96.6	205	21	Streptococcus pneu
8	687	96.6	489	21	Streptococcus pneu
9	687	96.6	780	21	Streptococcus pneu
10	687	96.6	840	21	Streptococcus pneu
11	687	96.6	1019	21	Streptococcus pneu

12	687	96.6	1019	21	Streptococcus pneu
13	687	96.6	1019	21	Streptococcus pneu
14	687	96.6	1019	21	Streptococcus pneu
15	687	96.6	1019	21	Streptococcus pneu
16	687	96.6	1019	21	Streptococcus pneu
17	687	96.6	1019	21	Streptococcus pneu
18	687	96.6	1019	21	Streptococcus pneu
19	605	85.1	840	21	Streptococcus pneu
20	600.5	84.5	827	21	Streptococcus pneu
21	599.5	84.3	826	21	Streptococcus pneu
22	599.5	84.3	838	21	Streptococcus pneu
23	599.5	84.3	838	21	Streptococcus pneu
24	596	83.8	334	21	Streptococcus pneu
25	596	83.8	690	21	Streptococcus pneu
26	596	83.8	821	21	Streptococcus pneu
27	596	83.8	821	21	Streptococcus pneu
28	592	83.3	819	21	Streptococcus pneu
29	589.5	82.9	819	21	Streptococcus pneu
30	585.5	82.3	807	21	Streptococcus pneu
31	585.5	82.3	811	21	Streptococcus pneu
32	585.5	82.3	819	21	Streptococcus pneu
33	585.5	82.3	819	21	Streptococcus pneu
34	585.5	82.3	820	21	Streptococcus pneu
35	585.5	82.3	834	21	Streptococcus pneu
36	581.5	81.8	805	21	Streptococcus pneu
37	581.5	81.8	811	21	Streptococcus pneu
38	581.5	81.8	811	21	Streptococcus pneu
39	581.5	81.8	816	21	Streptococcus pneu
40	581.5	81.8	816	21	Streptococcus pneu
41	581.5	81.8	816	21	Streptococcus pneu
42	575.5	80.9	763	19	Streptococcus pneu
43	574.5	80.8	811	21	Streptococcus pneu
44	574	80.7	796	19	Streptococcus pneu
45	428	60.2	825	22	Streptococcus pneu

ALIGNMENTS

RESULT. 1

AB12718
ID AAB12718 standard; Protein; 484 AA.

AC AAB12718;
DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-3A protein antigen SEQ ID NO:8.

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

PN WO200039299-A2.

PD 06-JUL-2000.

PF 20-DEC-1999; 99WO-CA01218.

PR 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

DR WPI; 2000-452397/39.

XX N-PSDB; AA65733.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
otitis media, bacteraemia and/or pneumonia

35 Claim 18; Fig 8; 106pp; English.
36
37 The present invention describes nucleic acids (I) encoding protein
38 antigens (II) from Streptococcus pneumoniae. The protein antigens
39 have bactericidal activity. The nucleic acids, encoding the proteins
40 antigens, may be used for the recombinant production of the proteins
41 they encode. The protein antigens may then be used as vaccines for the
42 prevention and treatment of Streptococcal infections in mammals
43 (especially humans) which result in, e.g. meningitis, otitis media,
44 bacteraemia and/or pneumonia. The present sequence represents the
45 S. pneumoniae BVH-3A protein antigen.
46
47 Sequence 484 AA:

Query Match 100.0%; Score 711; DB 21; Length 484;
Best Local Similarity 100.0%; Pred. No. 8.3e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IVSLSLCAVYALNQHSQENKNNRVSYVDGSSQSSQKSENLPDQVSOREGIOAEQIVTKI 60
15 IVSLSLCAVYALNQHSQENKNNRVSYVDGSSQSSQKSENLPDQVSOREGIOAEQIVTKI 74
61 TDGQVYVSHGDIHYHNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVQDKY 120
75 TDGQVYVSHGDIHYHNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVQDKY 134
121 VYVLEKDAHADNVRTK 136
135 VYVLEKDAHADNVRTK 150

RESULT 2
AB01467 standard; Protein: 484 AA.

AB01467;
20-OCT-2000 (first entry)

Recombinant variant of Sp36 (Sp36E) of S. pneumoniae.

Streptococcus pneumoniae; infection; vaccine; collid coll region;
histidine triad residue; Sp36; antibody; otitis media;
nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
meningitis; lobar pneumonia.

Streptococcus pneumoniae.

Key Location/Qualifiers
Region 63..68 /label- Histidine triad residue
Region 120..140 /label- Collid coll region
Region 185..190 /label- Histidine triad residue
Region 289..294 /label- Histidine triad residue
Region 376..381 /label- Histidine triad residue
Region 441..446 /label- Histidine triad residue
Region 750..772 /label- Histidine triad residue
Region /label- Collid coll region

W0200037105-A2.

29-JUN-2000.

21-DEC-1999; 99MO-US03090.

21-DEC-1998; 98US-0113048.

PA (MED-) MEDIMUNE INC.
XX Johnson LS, Koenig S, Adamou JE;
XX WPI: 2000-452129/39.
DR N-PSDB: AAA47603.
XX
XX Vaccine useful for prophylaxis and treatment of pneumococcal infections
XX such as otitis media, nasopharyngeal and bronchial infections,
XX comprises Streptococcus pneumoniae proteins
XX
XX Claim 1; Page 58-60; 70pp; English.

Although a number of proteins have been suggested as being involved
in the pathogenicity of Streptococcus pneumoniae, there still remains
a need to identify polypeptides having epitopes in common from
various strains of S. pneumoniae in order to utilize such
polypeptides in vaccines to protect against a wide variety of
S. pneumoniae. New vaccine compositions are described which comprise a
Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
acids in length that comprise at least one histidine triad residue
(HXHXH) or a collid-coll region, or an antibody directed against
these features. The vaccine is useful in protecting against infection
by Streptococcus pneumoniae. The vaccine composition comprising
antibodies to is useful for passive immunization for treating
pneumococcal infections which includes otitis media, nasopharyngeal
and bronchial infections.

Sequence 484 AA:

Query Match 100.0%; Score 711; DB 21; Length 484;
Best Local Similarity 100.0%; Pred. No. 8.3e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IVSLSLCAVYALNQHSQENKNNRVSYVDGSSQSSQKSENLPDQVSOREGIOAEQIVTKI 60
15 IVSLSLCAVYALNQHSQENKNNRVSYVDGSSQSSQKSENLPDQVSOREGIOAEQIVTKI 74
61 TDGQVYVSHGDIHYHNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVQDKY 120
75 TDGQVYVSHGDIHYHNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVQDKY 134
121 VYVLEKDAHADNVRTK 136
135 VYVLEKDAHADNVRTK 150

RESULT 3

AA01708 standard; Protein: 484 AA.

AA01708;

02-JUN-2000 (first entry)

Streptococcus pneumoniae protein sequence ID128.

Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
kidney disease; diabetes; immunosuppressive disorder; otitis media;
pneumococcal septicaemia; sinusitis; meningitis; therapy.

Streptococcus pneumoniae.

W0200006738-A2.

10-FEB-2000.

27-JUL-1999; 99MO-GB02452.

27-JUL-1998; 98GB-0016336.

19-MAR-1999; 99US-0125329.

XX (MICH-) MICROBIAL TECHNIQS LTD.
XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
XX WPI: 2000-195301/17.
XX N-PSDB: AA291804.

XX Streptococcal proteins and polynucleotides useful for diagnosis,
XX treatment and prophylaxis of bacterial infections

PS Claim 1: Page 39; 76pp; English.

XX This sequence represents a Streptococcus pneumoniae protein of the
XX invention. The proteins (or their homologues, derivatives and/or
XX fragments) are useful as immunogens or antigens. Immunogenic or antigenic
XX compositions comprising the proteins are useful as vaccines and also in
XX diagnostic assays. The sequences are useful for the detection or
XX diagnosis of S. pneumoniae infection, by contacting a sample to be tested
XX with them. Agents capable of antagonising, inhibiting or interfering with
XX the function or expression of the protein or polypeptide are useful in
XX medical compositions in the treatment or prophylaxis of S. pneumoniae
XX infection. As the sequences can be used to treat S. pneumoniae infection,
XX they can be used to treat bacterial pneumonia, which has high rates in
XX young children, the elderly, and in patients with predisposing conditions
XX such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
XX or with immunosuppressive disorders, especially AIDS. They can also be
XX used to treat pneumococcal septicaemia, otitis media, sinusitis, and
XX meningitis.

XX Sequence 484 AA:

Query Match 100.0%; Score 711; DB 21; Length 484;
Best Local Similarity 100.0%; Pred. No. 8.3e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 IYSLISCAVALNQHSEKDKNNRVSYVDGSSQSKSENLTPDOVSQKGIQAEQIVIRI 60
15 IYSLISCAVALNQHSEKDKNNRVSYVDGSSQSKSENLTPDOVSQKGIQAEQIVIRI 74
61 TQGGYVTSQGDHRYNGKVPYDALFSEELMKDPYQKADYNEVAGGVIIRYDGRY 120
75 TGGYVTSQGDHRYNGKVPYDALFSEELMKDPYQKADYNEVAGGVIIRYDGRY 134
QY 121 YYYLKDAHADNVRTK 136
135 YYYLKDAHADNVRTK 150
DB 135 YYYLKDAHADNVRTK 150

RESULT 4
ID AAY81538 standard; Protein; 485 AA.

XX AAY81538;

XX 24-MAY-2000 (first entry)

XX Streptococcus pneumoniae type 4 protein sequence #38.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;
XX antibacterial; antiinflammatory; meningitis; infection; diagnosis;
XX pneumococcal disease.

XX Streptococcus pneumoniae.

XX WO200006737-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02451.

XX 27-JUL-1998; 98GB-0016337.

PR 19-MAR-1999; 99US-0125164.
XX (MICH-) MICROBIAL TECHNIQS LTD.
XX Gilbert CFG, Hansbro PM;
XX WPI: 2000-195300/17.

XX New Streptococcal protein, useful as a vaccine, for diagnosis of
XX pneumococcal diseases and for screening agents capable of antagonizing
XX or inhibiting expression of the protein

PS Claim 1: Page 76; 108pp; English.

XX AA05101 to AA05167 represent specifically claimed protein sequences
XX isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent
XX specifically claimed nucleotide sequences isolated from S. pneumoniae.
XX The sequences have antibacterial and antiinflammatory properties.
XX The protein sequences, and fragments of them, are useful as immunogens
XX and/or antigens. The nucleotide sequences can be used in vaccines and in
XX diagnostic assays. The proteins and nucleotides can be useful for the
XX detection and diagnosis of S. pneumoniae. The protein sequences are also
XX useful for screening an agent capable of antagonising, inhibiting or
XX interfering with the function or expression of the proteins in which the
XX agent is useful for treatment or prophylaxis of S. pneumoniae infection
XX and meningitis. AA05591 to AA05614 represent primers used in the
XX exemplification of the present invention.

XX Sequence 485 AA:

Query Match 100.0%; Score 711; DB 21; Length 485;
Best Local Similarity 100.0%; Pred. No. 8.3e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 IYSLISCAVALNQHSEKDKNNRVSYVDGSSQSKSENLTPDOVSQKGIQAEQIVIRI 60
15 IYSLISCAVALNQHSEKDKNNRVSYVDGSSQSKSENLTPDOVSQKGIQAEQIVIRI 74
61 TQGGYVTSQGDHRYNGKVPYDALFSEELMKDPYQKADYNEVAGGVIIRYDGRY 120
75 TGGYVTSQGDHRYNGKVPYDALFSEELMKDPYQKADYNEVAGGVIIRYDGRY 134
QY 121 YYYLKDAHADNVRTK 136
135 YYYLKDAHADNVRTK 150
DB 135 YYYLKDAHADNVRTK 150

RESULT 5
ID AAB12724 standard; Protein; 509 AA.

XX AAB12724;

XX 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae L-BVH-3-AD protein antigen SEQ ID NO:57.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
XX otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

XX WO200039299-A2.

XX 06-JUL-2000.

XX 20-DEC-1999; 99WO-CA01218.

XX 23-DEC-1998; 98US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.

XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI: 2000-452397/39.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteremia and/or pneumonia
XX
XX Claim 18; Fig 22; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteremia and/or pneumonia. The present sequence represents the
XX S. pneumoniae L-BVH-3 AD protein antigen.

XX Sequence 509 AA;

Query Match 100.0%; Score 711; DB 21; Length 509;
Best Local Similarity 100.0%; Pred. No. 8.9e-70;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IVSLCAVYALNQRSEKNDNRVSYVDSQSSQSKSENTPPDVQSGKEGIQADQIVIKI 60
15 IVSLSICAYALNQRSEKNDNRVSYVDSQSSQSKSENTPPDVQSGKEGIQADQIVIKI 74
61 TDGQVYVSHGDHYHYNGKVPYDALFSEELMKDPNQQLKADIVNEVKGYYIIRVDKXY 120
75 TDGQVYVSHGDHYHYNGKVPYDALFSEELMKDPNQQLKADIVNEVKGYYIIRVDKXY 134
121 YVYIKDAAHADNVTK 136
135 YVYIKDAAHADNVTK 150

RESULT 6
AAB12715 standard; Protein; 1039 AA.

XX AAB12715;
XX 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae BVH-3 protein antigen SEQ ID NO:2.
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
XX otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

WC0200039299-A2.

06-JUL-2000.

20-DEC-1999; 99MO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI: 2000-452397/39.

N-PSDB; AAA65730.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteremia and/or pneumonia -

XX Claim 18; Fig 2; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteremia and/or pneumonia. The present sequence represents the
XX S. pneumoniae BVH-3 protein antigen.

XX Sequence 1039 AA;

Query Match 100.0%; Score 711; DB 21; Length 1039;
Best Local Similarity 100.0%; Pred. No. 2.4e-69;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IVSLCAVYALNQRSEKNDNRVSYVDSQSSQSKSENTPPDVQSGKEGIQADQIVIKI 60
15 IVSLSICAYALNQRSEKNDNRVSYVDSQSSQSKSENTPPDVQSGKEGIQADQIVIKI 74
61 TDGQVYVSHGDHYHYNGKVPYDALFSEELMKDPNQQLKADIVNEVKGYYIIRVDKXY 120
75 TDGQVYVSHGDHYHYNGKVPYDALFSEELMKDPNQQLKADIVNEVKGYYIIRVDKXY 134
121 YVYIKDAAHADNVTK 136
135 YVYIKDAAHADNVTK 150

RESULT 7
AAB12726 standard; Protein; 205 AA.

XX AAB12726;
XX 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae BVH-3C protein antigen SEQ ID NO:59.
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteremia;
XX otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

WC0200039299-A2.

06-JUL-2000.

20-DEC-1999; 99MO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI: 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteremia and/or pneumonia -

Claim 18; Fig 24; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the proteins
XX they encode, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the

prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents the S. pneumoniae BVH-3C protein antigen.

Sequence 205 AA;

Query Match 96.6%; Score 687; DB 21; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.1e-67;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNQHRSQENKNNRRSYVDGSSQSKSENLTDPQVSQKKGIOAEQIVIKITDQGYV 66
1 cayalnghrsgeknknrrsyvdsqsgsksenlt dpqvsqkkgiaeqivikltdqgyv 60
QY 67 TSHGDHYHYNGKVPYDALFSEELMKDPYOLKADIVNEVGKGIIRKVDGKYYVYLKD 126
61 tshgdhyhyngkvrpydal fseellmkdpnyqlkdadivnevsgylikvdgkyyyylkd 120
QY 127 AAADNVKTK 136
121 aaadnvrtk 130
Db

RESULT 8

AAB12723 8
ID AAB12723 standard; Protein; 489 AA.

AAB12723;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-3AD protein antigen SEQ ID NO:56.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

XX MO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA01218.

23-DEC-1996; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
OTitis media, bacteraemia and/or pneumonia -

PS Claim 18; Fig 21; 106pp; English.

CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3AD protein antigen.

XX Sequence 489 AA;

Query Match 96.6%; Score 687; DB 21; Length 489;
Best Local Similarity 100.0%; Pred. No. 3.7e-67;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNQHRSQENKNNRRSYVDGSSQSKSENLTDPQVSQKKGIOAEQIVIKITDQGYV 66
1 cayalnghrsgeknknrrsyvdsqsgsksenlt dpqvsqkkgiaeqivikltdqgyv 60
QY 67 TSHGDHYHYNGKVPYDALFSEELMKDPYOLKADIVNEVGKGIIRKVDGKYYVYLKD 126
61 tshgdhyhyngkvrpydal fseellmkdpnyqlkdadivnevsgylikvdgkyyyylkd 120
QY 127 AAADNVKTK 136
121 aaadnvrtk 130
Db

RESULT 9

AAB12744 9
ID AAB12744 standard; Protein; 780 AA.

AAB12744;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae NEW15 protein antigen SEQ ID NO:78.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

XX MO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA01218.

23-DEC-1996; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
OTitis media, bacteraemia and/or pneumonia -

PS Claim 18; Fig 43; 106pp; English.

CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae NEW15 protein antigen.

XX Sequence 780 AA;

Query Match 96.6%; Score 687; DB 21; Length 780;
Best Local Similarity 100.0%; Pred. No. 7.1e-67;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNQHRSQENKNNRRSYVDGSSQSKSENLTDPQVSQKKGIOAEQIVIKITDQGYV 66
1 cayalnghrsgeknknrrsyvdsqsgsksenlt dpqvsqkkgiaeqivikltdqgyv 60
Db

QY 67 TSHGHHYHNGKVPYDALLFSEELLMKDPNTQKADIVNEKGGYIIKVDGKYYVYLKD 126
DB 61 tshgddhyhngkvpdydallfseellmkdpnyqlkdadivnevkggyllkvdgkyyvylkd 120
QY 127 AAHADNVTRK 136
DB 121 aahadnvtrk 130

RESULT 10
AAB12721

ID AAB12721 standard; Protein; 840 AA.

AC AAB12721;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae Sp63 BVH-3 protein antigen SEQ ID NO:16.

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

MO200039299-A2.

06-JUL-2000.

PF 20-DEC-1999; 99MO-CA01218.

PR 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Plneau I, Martin D, Rioux C, Charland N;

DR WPI: 2000-452397/39.

DR N-PSDB; AAA65738.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia

PS Claim 18; Fig 19; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents the
XX S. pneumoniae Sp63 BVH-3 protein antigen.

Sequence 840 AA;

Query Match 96.6%; Score 687; DB 21; Length 840;
Best Local Similarity 100.0%; Pred. No. 7.9e-67;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNHRSGENKDNRRVSYVDSOSSQKSENLRPPQVDSQREGIOAEQIVIKITDGGYV 66
DB 1 cayalnhrsgenkdnnrvsyvdsqsgksenlrppdqvsqegivikltdggyv 60

QY 67 TSHGHHYHNGKVPYDALLFSEELLMKDPNTQKADIVNEKGGYIIKVDGKYYVYLKD 126
DB 61 tshgddhyhngkvpdydallfseellmkdpnyqlkdadivnevkggyllkvdgkyyvylkd 120

QY 127 AAHADNVTRK 136

DB 121 aahadnvtrk 130

RESULT 11
AAB12722
ID AAB12722 standard; Protein; 1019 AA.
AC AAB12722;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

MO200039299-A2.

06-JUL-2000.

PF 20-DEC-1999; 99MO-CA01218.

PR 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Plneau I, Martin D, Rioux C, Charland N;

DR WPI: 2000-452397/39.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia

PS Claim 18; Fig 20; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the proteins
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents the
XX S. pneumoniae BVH-3M protein antigen.

Sequence 1019 AA;

Query Match 96.6%; Score 687; DB 21; Length 1019;
Best Local Similarity 100.0%; Pred. No. 1e-66;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNHRSGENKDNRRVSYVDSOSSQKSENLRPPQVDSQREGIOAEQIVIKITDGGYV 66
DB 1 cayalnhrsgenkdnnrvsyvdsqsgksenlrppdqvsqegivikltdggyv 60

QY 67 TSHGHHYHNGKVPYDALLFSEELLMKDPNTQKADIVNEKGGYIIKVDGKYYVYLKD 126
DB 61 tshgddhyhngkvpdydallfseellmkdpnyqlkdadivnevkggyllkvdgkyyvylkd 120

QY 127 AAHADNVTRK 136

DB 121 aahadnvtrk 130

RESULT 12

ID AAB12748 standard; Protein; 1019 AA.

AC AAB12748;

XX AAB12748;

21-NOV-2000 (first entry)
Streptococcus pneumoniae strain WU2 BVH-3 protein antigen.
Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
otitis media; pneumonia; immunisation; bactericidal.
Streptococcus pneumoniae.
MO200039299-A2.
06-JUL-2000.
20-DEC-1999; 99NO-CA01218.
23-DEC-1998; 98US-0113800.
(BIOC-) BIOCHEM PHARMA INC.
Hamel J, Brodeur BR, Plneau I, Martin D, Rioux C, Charland N;
MPI; 2000-452397/39.
Streptococcal antigens useful for vaccinating against e.g. meningitis,
otitis media, bacteraemia and/or pneumonia -
Disclosure; Fig 11; 106pp; English.
The present invention describes nucleic acids (I) encoding protein
antigens (II) from Streptococcus pneumoniae. The protein antigens
have bactericidal activity. The nucleic acids, encoding the proteins
antigens, may be used for the recombinant production of the proteins
they encode. The protein antigens may then be used as vaccines for the
prevention and treatment of Streptococcal infections in mammals
(especially humans) which result in, e.g. meningitis, otitis media,
bacteraemia and/or pneumonia. The present sequence represents a
S. pneumoniae BVH-3 protein antigen, from the present invention.
Sequence 1019 AA;

Query Match 96.6%; Score 687; DB 21; Length 1019;
Best Local Similarity 100.0%; Pred. No. 1e-66;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

7 CAYALNHRSGENKDNKRVSYVDGSSQSKSENTLPDQVSGKEIGIAEQIYIKITDGYV 66
1 cayalnhrsgenkdnrvsyvdgssqsksentlpdqvsgkeigieqiyikitdgyv 60
67 TSHGDHYHYNGKVPYALPSEELMKDPNYQLKADIVNEVGKGIITKVGKTYVTLKD 126
61 tshgdhyhyngkvpdalpseellmkdpnyqlkadivnevgkgylltkvgktyvylkd 120
127 AAHADNVRTK 136
121 aahadnvr tk 130

RESULT 13
AA012749
AA012749 standard; Protein; 1019 AA.

21-NOV-2000 (first entry)
Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.
Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.
MO200039299-A2.
06-JUL-2000.
20-DEC-1999; 99NO-CA01218.
23-DEC-1998; 98US-0113800.
(BIOC-) BIOCHEM PHARMA INC.
Hamel J, Brodeur BR, Plneau I, Martin D, Rioux C, Charland N;
MPI; 2000-452397/39.
Streptococcal antigens useful for vaccinating against e.g. meningitis,
otitis media, bacteraemia and/or pneumonia -
Disclosure; Fig 11; 106pp; English.
The present invention describes nucleic acids (I) encoding protein
antigens (II) from Streptococcus pneumoniae. The protein antigens
have bactericidal activity. The nucleic acids, encoding the proteins
antigens, may be used for the recombinant production of the proteins
they encode. The protein antigens may then be used as vaccines for the
prevention and treatment of Streptococcal infections in mammals
(especially humans) which result in, e.g. meningitis, otitis media,
bacteraemia and/or pneumonia. The present sequence represents a
S. pneumoniae BVH-3 protein antigen, from the present invention.
Sequence 1019 AA;

Query Match 96.6%; Score 687; DB 21; Length 1019;
Best Local Similarity 100.0%; Pred. No. 1e-66;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

7 CAYALNHRSGENKDNKRVSYVDGSSQSKSENTLPDQVSGKEIGIAEQIYIKITDGYV 66
1 cayalnhrsgenkdnrvsyvdgssqsksentlpdqvsgkeigieqiyikitdgyv 60
67 TSHGDHYHYNGKVPYALPSEELMKDPNYQLKADIVNEVGKGIITKVGKTYVTLKD 126
61 tshgdhyhyngkvpdalpseellmkdpnyqlkadivnevgkgylltkvgktyvylkd 120
127 AAHADNVRTK 136
121 aahadnvr tk 130

RESULT 14
AA012750
AA012750 standard; Protein; 1019 AA.

21-NOV-2000 (first entry)
Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.
Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.
MO200039299-A2.
06-JUL-2000.
20-DEC-1999; 99NO-CA01218.

PR 23-DEC-1998; 9805-0113800.
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 OR WPI; 2000-452397/39.
 XX
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 PS Disclosure; Fig 11; 106pp; English.
 CC
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins.
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 XX
 SO Sequence 1019 AA;

Query Match 96.6%; Score 687; DB 21; Length 1019;
 Best Local Similarity 100.0%; Pred. No. 1e-66;
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 7 CAYALNQHRSQENKNNRRSYVDGSSQSKSENLTPDOVSQKGIQAEQIVYKITDQGYV 66
 |||||||
 Db 1 cayalnqhrrsgenknrrsyvdsqsgsksenltpdqvsqkglqaeqlvkltdqgyv 60
 67 TSHGDHYHYNGKVPYDALFSEELMKDPNYOLKADIVNEVGIIIVDKGYYYLND 126
 |||||||
 Db 61 tshgdhyhyngkvpdydalffseellmkdpnyqlkdadivnevgiylkvdgkyyyylnd 120
 2Y 127 AAHADNVRTK 136
 |||||||
 Db 121 aahadvrtk 130

RESULT 15
 AAB12751
 ID AAB12751 standard; Protein; 1019 AA.
 NC
 AAAB12751;
 XX
 XX 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae strain SP64 BVH-3 protein antigen.
 XX
 XX Streptococcus pneumoniae: BVH-3; BVH-11; BVH-28; antigen; vaccine;
 XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 XX otitis media; pneumonia; immunisation; bactericidal.
 XX

XX Streptococcus pneumoniae.
 XX
 XX MO200039299-AA2.
 XX
 XX 06-JUL-2000.
 XX
 XX 20-DEC-1999; 99WO-CA01218.
 XX
 XX 23-DEC-1998; 9805-0113800.
 XX
 XX (BIOC-) BIOCHEM PHARMA INC.
 XX
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 XX WPI; 2000-452397/39.
 XX

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 XX
 PS Disclosure; Fig 11; 106pp; English.
 CC
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 XX
 SO Sequence 1019 AA;

Query Match 96.6%; Score 687; DB 21; Length 1019;
 Best Local Similarity 100.0%; Pred. No. 1e-66;
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAYALNQHRSQENKNNRRSYVDGSSQSKSENLTPDOVSQKGIQAEQIVYKITDQGYV 66
 |||||||
 Db 1 cayalnqhrrsgenknrrsyvdsqsgsksenltpdqvsqkglqaeqlvkltdqgyv 60
 QY 67 TSHGDHYHYNGKVPYDALFSEELMKDPNYOLKADIVNEVGIIIVDKGYYYLND 126
 |||||||
 Db 61 tshgdhyhyngkvpdydalffseellmkdpnyqlkdadivnevgiylkvdgkyyyylnd 120
 QY 127 AAHADNVRTK 136
 |||||||
 Db 121 aahadvrtk 130

Search completed: September 26, 2001, 22:30:30
 Job time: 3209 sec

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED
DATE 10/05/01 BY 60324 UCBAW

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

protein - protein search, using sw model

Run on: September 26, 2001, 22:33:45 ; Search time 73.11 Seconds
(without alignments)
246.115 Million cell updates/sec

Title: US-09-471-255-2_COPY_15_150

Perfect score: 711
Sequence: 1 IVSLSLCAVALNQHRSQENK.....DKYTYVLKDAHADNVRK 136

oring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SPTREMBL_16:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mmc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.unclassified:*
- 13: sp.vertebrate:*
- 14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	423	59.5	822	2	Q9ZHG7 streptococ
2	103	14.5	1089	10	Q9FN97 arabidopsi
3	88	12.4	198	2	053037 haemophil
4	87.5	12.3	736	13	093563 rana ridib
5	85.5	12.0	721	2	09PR88 ureaplasma
6	84.5	11.9	1877	2	09XXM1 plasmodium
7	83	11.7	814	2	09PHD9 xylella fas
8	82	11.5	1946	2	048545 lactobacilli
9	81.5	11.5	703	5	046944 aplysia cal
10	81	11.4	712	5	016971 aplysia cal
11	79.5	11.2	567	2	09ZM97 helicobacte
12	79.5	11.2	2573	5	096185 plasmodium
13	78	11.0	264	2	095337 mycoplasma
14	78	11.0	438	2	046298 clostridium
15	78	11.0	642	2	09PHM6 campylobac
16	77.5	10.9	206	3	074190 kluyveromyc
17	77.5	10.9	857	10	09T0C8 arabidopsi
18	77	10.8	776	5	015793 p. strain dd
19	77	10.8	1766	5	Q25668 plasmodium

20	77	10.8	1785	5	Q25685	Q25685 plasmodium
21	76.5	10.8	490	2	Q9RMV1	Q9RMV1 bacillus an
22	76.5	10.8	1799	5	Q18220	Q18220 caenorhabd1
23	76.5	10.8	1802	5	Q18219	Q18219 caenorhabd1
24	76	10.7	608	10	Q9FFG6	Q9FFG6 arabidopsi
25	76	10.7	875	5	Q9P7S5	Q9P7S5 schizosach
26	76	10.7	2421	5	Q9VM48	Q9VM48 drosophila
27	75.5	10.6	312	2	050603	050603 staphylococ
28	75.5	10.6	332	9	080066	080066 bacterioph
29	75.5	10.6	472	5	Q9MOB6	Q9MOB6 drosophila
30	75	10.5	214	5	Q18307	Q18307 caenorhabd1
31	75	10.5	1060	5	Q16926	Q16926 aedes aegypt
32	74.5	10.5	388	1	Q28907	Q28907 archaeglob
33	74.5	10.5	753	6	Q9GLR1	Q9GLR1 bos taurus
34	74	10.4	710	4	Q9UM02	Q9UM02 homo sapien
35	74	10.4	802	5	Q9V773	Q9V773 drosophila
36	73.5	10.3	472	5	Q9GR08	Q9GR08 plasmodium
37	73.5	10.3	1007	5	Q96848	Q96848 dictyostell
38	73	10.3	282	2	050852	050852 borrelia bu
39	73	10.3	506	4	Q14898	Q14898 homo sapien
40	73	10.3	556	2	Q9XW3	Q9XW3 porphyromon
41	73	10.3	356	2	Q9R0J2	Q9R0J2 porphyromon
42	73	10.3	710	11	Q70196	Q70196 rattus norv
43	73	10.3	388	1	Q9QUR6	Q9QUR6 mus musculu
44	73	10.3	1590	2	Q59983	Q59983 streptococ
45	72.5	10.2	385	14	Q9JFJ1	Q9JFJ1 helicoverpa

ALIGNMENTS

RESULT	1	PREDIMINARY:	PRT:	822 AA.
Q9ZHG7	Q9ZHG7			
AC	Q9ZHG7			
DT	01-MAY-1999 (TREMUREL. 10, Created)			
DT	01-MAY-1999 (TREMUREL. 10, Last sequence update)			
DT	01-MAY-2000 (TREMUREL. 13, Last annotation update)			
DE	HYPOTHETICAL. 92.4 KDA PROTEIN.			
OS	Streptococcus agalactiae.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
CC	Streptococcus.			
OX	NCBI_TaxID-1311;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-R268;			
RX	MEDLINE-99115568; PubMed-9916102;			
RA	Spellenberg B., Rozdzinski E., Martin S., Weber-Heymann J.,			
RT	Schmitzler N., Luetticken R., Podbielski A.;			
RT	"Lmb, a protein with similarities to the Lrai adhesin family, mediates			
RL	attachment of Streptococcus agalactiae to human laminin."			
DR	EMBL; AF062533; AAD13797.1;			
KW	Hypothetical protein.			
Q9ZHG7	SEQUENCE 822 AA; 92385 MW; 80E4EDF313481F98 CRC64;			
Query Match	59.5%; Score 423; DB 2; Length 822;			
Best Local Similarity	55.1%; Pred. No. 4.2e-29;			
Matches	76; Conservative 26; Mismatches 34; Indels 2; Gaps 1;			
QY	1 IVSLSLCAVALNQHRSQENKNNRVSYYDGSQSSQKS--ENTTPQVQSKGIOAEQIV 58			
DB	15 LLAFTIGSYOUGKHMGLATKDNQIAYTIDDSGKVKAKTKNTMDQISAEBISAEQIV 74			
QY	59 KITDGGYTSQGDHYHYNGKVPYDALFSEELAMDPNYQLKADIVAEVGGYLYIKNG 118			
DB	75 KITDGGYTSQGDHYHYNGKVPYDALFSEELAMDPNYHYHQSVDVINEILDGYIYKNG 134			
QY	119 KYTYVLKDAHADNVRK 136			
DB	135 NYTYVLKSGSKKNIRTK 152			

```

RESULT 2
ID 09FN97 PRELIMINARY; PRT: 1089 AA.
C 09FN97;
T 01-MAR-2001 (TREMblrel. 16, Created)
T 01-MAR-2001 (TREMblrel. 16, Last sequence update)
T 01-MAR-2001 (TREMblrel. 16, Last annotation update)
E TRANSPOSON PROTEIN-LIKE.
C Arabidopsis thaliana (Mouse-ear cross).
C Eukaryote: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
C Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
C Brassicales; Brassicaceae; Arabidopsis.
C NCBI_TaxID=3702;
N [1]
P SEQUENCE FROM N.A.
C STRAIN-COLDBIA;
X MEDLINE-98069011; PubMed-9405937;
A Kotani H., Nakamura T., Sato S., Kaneko T., Asamizu E., Miyajima N.,
A Tabata S.;
T Structural analysis of Arabidopsis thaliana chromosome 5. II.
T Sequence features of the regions of 1,044,062 bp covered by thirteen
T physically assigned pl clones.
T DNA Res. 4:291-300(1997).
L EMBL: AB006705; BAB09502.1;
P SEQUENCE 1089 AA; 126629 MW; 17664F2489EB14E CRC64;

Query Match 14.5%; Score 103; DB 10; Length 1089;
Best Local Similarity 23.7%; Pred. No. 0.76;
Matches 31; Conservative 25; Mismatches 73; Indels 2; Gaps 2;

Y 5 SICAYALNHRSDKNNRVSQSSQSKSENTLPDQVSQKEGIAEQIVIKITDQ 64
b 792 TFCYYGDBHKTNRAGDR HYDGNEDHEHDYDPIFSQAGRSKSEWLDDKD 850
Y 65 YVSHGHHYNGKVPYDALFSEELMKDPNOLKADIVNEVK-GGYIIKVDGKYYV 123
b 851 YHMHRYTLRNCDDLRFERFLFDESLIANPDSKDLNELREKQSSWLKYNSTPEW 910
Y 124 LKDAAHADNVR 134
b 911 LLSIVHGPMVK 921

ESTUT 3
ID 053037 PRELIMINARY; PRT: 198 AA.
C 053037;
T 01-JUN-1998 (TREMblrel. 06, Created)
T 01-JUN-1998 (TREMblrel. 06, Last sequence update)
T 01-NOV-1998 (TREMblrel. 08, Last annotation update)
E PUTATIVE HAEMOCIN PROCESSING PROTEIN.
C HMCC.
C Haemophilus influenzae.
C Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
C Haemophilus.
C NCBI_TaxID=727;
N [1]
P SEQUENCE FROM N.A.
C STRAIN-ELA;
X MEDLINE-97197533; PubMed-9045829;
A Murley Y.M., Edling T.D., Pozsgay J.M., Lipuma J.J.;
A Cloning and characterization of the haemocin immunity gene of
A Haemophilus influenzae.
A J. Bacteriol. 179:1684-1689(1997).
N [2]
P SEQUENCE FROM N.A.
C STRAIN-ELA;
A Murley Y.M., Edling T.D., Pozsgay J.M., Lipuma J.J.;
A Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
N [3]
P SEQUENCE FROM N.A.

```

```

RC STRAIN-ELA;
RA Murley Y.M., Edling T.D., Pozsgay J.M., Lipuma J.J.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: U68399; AAC46223.1;
SQ SEQUENCE 198 AA; 22978 MW; 4C3F3F36A8723C95/CRC64;

Query Match 12.4%; Score 88; DB 2; Length 198;
Best Local Similarity 25.9%; Pred. No. 2;
Matches 39; Conservative 28; Mismatches 39; Indels 42; Gaps 9;

Y 14 HRSQKND-----NNRVSQSSQSK-SENTLPDQVSQKEGIAEQIVIKITDQ 63
b 63 HISEINEDTILYKLNNEEASFELARISKEFNINAGLTLTKELINIKPVI----- 117
Y 64 GYVS--HGHHYNGKVPYDALFSEELMKDP--NYQAKD-----YVNEVGGY 112
b 118 AYVNSLSNNDHFIING-----IFNKELLSDPAIGNYSLSKSDFERIMLNRDKKGI 171
Y 113 IIRYDGKYYVYLKDAAH---ADNVRTK 136
b 172 L-----YLRSDKHLEFDINIKT 191

RESULT 4
ID 093583 PRELIMINARY; PRT: 736 AA.
AC 093583;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE PHOROMONE CONVERTASE 1 PRECURSOR.
OS Rana ridibunda (laughing frog) (Marsh frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
OX NCBI_TaxID=8406;
RN [1]
RP SEQUENCE FROM N.A.
RA Gangnon F., Jangou S., Vileau D., Seidah N.G., Vaudry H.;
RT Cloning of frog PC1.
RL J. Comp. Neurol. 0:0-0(1998).
DR EMBL: AF092904; AAC62483.1;
DR InterPro: IPR002884;
DR Pfam: PF00082; Peptidase_S8; 1.
DR Pfam: PF01483; P. 1.
DR PRINTS: PRO0723; SUBTILISIN.
DR PRODOM: PD000717; -; 1.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 736 AA; 82965 MW; 84C00B44EB8C3CD7 CRC64;

Query Match 12.3%; Score 87.5; DB 13; Length 736;
Best Local Similarity 23.2%; Pred. No. 11;
Matches 36; Conservative 20; Mismatches 44; Indels 55; Gaps 5;

Y 5 SICAYALNHRSDKND-----DNNRVSQSSQSK-----SENT 41
b 63 SLGTHLFLRHRDQPRRSRSAPATIRKLVDNRSWAEOOYIQQFKRGYVYANTDSEDLF 122
Y 42 PDQVSQ-----REGIAEQIVIKITDQGVTSFGDHYHY 76
b 123 NDKVPMKNQVTLNDRVNPPLPKLIDLVIVYVWKRGLTGKGSVTVLDDGLENNHTDITANT 182
Y 77 NGKVPYDALFSEELMKDPNOLKADIVNEVGG 111
b 183 DPEASVD--FNDNDKPPRY-----DITENKKG 210

RESULT 5
ID 09PR88

```

```

01 09PR88; PRELIMINARY; PRT; 721 AA.
AC 09PR88;
CT 01-OCT-2000 (TEMBLrel. 15, Created)
CT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
CT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
3E VIRULENCE ASSOCIATED PROTEIN-EXORIBONDLEASES.
3N VAB.
7S Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
2C Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscates;
3C Mycoplasmataceae; Ureaplasma.
CX NCBI_TaxID=134821;
CX [1]
RP SEQUENCE FROM N.A.
RP SPRAIN-SEROVAR 3;
5C MEDLINE=20500219; PubMed=11048724;
RA Glass J.T., Lefkowitz E.C., Glass J.S., Heiner C.R., Chen E.Y.
RT Cassell G.H.;
RT "the complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum";
RT Nature 407:757-762(2000).
RL EMBL: AE002105; AF30462.1; -
DR InterPro; IPR001900; -
DR InterPro; IPR003029; -
DR Pfam; PF00575; S1; 1
DR Pfam; PF00773; RNB; 1
DR PROSITE; PS01175; RBOUONDLEASE_II; 1
50 SEQUENCE 721 AA; 83173 MW; 67EEFA94AE4FD4AA CRC64;

```

```

Query Match      13.0%; Score 85.5; DB 2; Length 721;
Best Local Similarity 25.7%; Pred. No. 16;
Matches 36; Conservative 20; Mismatches 55; Indels 29; Gaps 7

QY 11 LNHRSQENKDNRR--VSIVDSQSQSKSENITPQVQSKEGIAQEQIYKTKDQGYVS 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 48 LIQENTIKLENNRPLVIGYLDYEFQHE-----IKGFI-----ITINSGDSFIKE 92
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 69 HGDNHYV-NGVKVPDALSSEL-----LMDPNYQAKMA---DIYNEVK-----GGYIKV 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 93 DNTLEIYYNNKYLINALKQDSYKFKVTKLEKPKNNLQDAVAEIVGHADHYGVGFIPLP 152
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

117 DGKVVVYLKDAAHADNVETK 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
153 NGGYIIFVDDEPLFIATINLK 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 6
 09XXM1
 ID 09XXM1 PRELIMINARY; PRT; 1877 AA.
 AC 09XXM1;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 P-TYPE ATPase.
 OS plasmodium yoelii.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 GX NCBI.TaxID=5861;
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC SPRAIN-17XL SPRAIN;
 RA Kimura M., Tanabe K., Krishna S., Tsudoi T., Saito-Ito A., Otani S.
 RA Ogura H.;
 RT "Characterization of a novel P-type ATPase in Plasmodium yoelii.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AB018798; BAA71326.1;
 DR InterPro: IPR001757;
 DR Pfam: PF000122; EI-E2_ATPase; 1.
 DR PRINTS: PR00119; CATPASE.
 DR PROSITE: PS00154; ATPASE_EI_E2; UNKNOWN_1.
 SO SEQUENCE 1877 AA; 220196 MW; 16974C388B01746E CRC64;

```

Best Local Similarity 21.8%; Pred. No. 62;
Matches 31; Conservative 35; Mismatches 63; Indels 13; Gaps 5

QY 1 IVS1:SCAYALNQH-RSOENKDNFNVSYSOQSSQSKSENLPD-QVSOKEIQAEQIYI 58
Db 848 INSTLEKNTINKNIREQNYNQTNRKLEIIDKSENSEKDFKSPNFSICKENSOSSNDLFF 907
QY 59 KITDOGYTTHSGDHHYHYNGK------PYALRSEELMLMDPNQLDLADIVNEVKG 111
Db 908 QSYDEGYOKSEHPPIRDEKITYTNHQSHPDLQKDEIRDAETR-KDEIRKDE-- 964
QY 112 YIKVYDGKYYVYLKDAADHNV 133
Db 965 -IRKNEENFNHSHNSLANSNNI 985

RESULT 7
Q9PHD9 PRELIMINARY; PRT: 814 AA.
AC Q9PHD9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE DNA GYRASE SUBUNIT B.
GN XE0005.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
CC Xylella.
CX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reilbach F.C., Arrida P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Azeia J.E., Baia G.S., Baptista C.S.,
RA Barros M.R., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
RA Coutinho N.B., Colombari C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Colantuoni L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferriz J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.S., Gomes S.L., Gruber A.,
RA Ho P.L., Honneisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Klieger J.E., Krummae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsushima A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Montello-Vitello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega-F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pequeiro J.B.,
RA Queglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silveira M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;
RL "The genome sequence of the plant pathogen xylella fastidiosa."
RT Nature 406:151-155(2000).
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA (BY SIMILARITY).
CC -1- SIMILARITY: TO DNA TOPOISOMERASE II FAMILY.
CC DB: EMBL: AE003855; AAF82818.1; -
DR InterPro: IPR000410; -
DR InterPro: IPR001241; -
DR InterPro: IPR002288; -
DR InterPro: IPR002936; -
DR InterPro: IPR003594; -
DR Pfam: PF00204; DNA_topoisomII_2.
DR Pfam: PF0086; DNA_gyraseB_C; 1.
DR Pfam: PF01751; TopoIIm; 1.

```

PRINTS: PR00418; TP12FAMILY
 PROSITE: PS00177; TOPOISOMERASE_II; 1.
 SMART: SMO0387; HATPase_C; 1.
 ATP-binding; Isomerase; Topoisomerase.
 SEQUENCE 814 AA; 90188 MW; 7C1EE231BC9A28A CRC64;

Query Match 11.7%; Score 83; DB 2; Length 814;
 Best Local Similarity 20.6%; Pred. No. 31;
 Matches 28; Conservative 21; Mismatches 45; Indels 42; Gaps 5;

27 YVDGSSQSKSENLPPDOYSGKEGIAQDIY-----KITDQGYYS--HG----- 70
 75 HVDGSVYSNDKRGIPVDINKEGSALEVITLVHAGKFDSDSKYSGGLHGVGSVV 134
 71 -----DHRYNGKVPYDALFSEELMKDPYOLKADIYNEVKGYYIKVD 117
 135 NALSRLMLDIWRDGYH-----QGEYVLGEPOYPLKOLGV--SAKRGTTLAFK 181
 118 GKYYVYLDMAHADNV 133
 182 PAKELFSDVEFHYENTL 197

RESULT 8
 048545 PRELIMINARY; PRT; 1946 AA.

01-NOV-1996 (TREMBLrel. 01, Created)
 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 PROTEINASE (PRT) PRECURSOR (PRTB).
 PRTB.
 S Lactococcus delbrueckii.
 C Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 Lactobacillus.
 X NCBI_Taxid=1584;

SEQUENCE FROM N.A.
 MEDLINE-96236017; PubMed-8655480;
 A Gilbert C., Atlan D., Blanc B., Portaller R., Germond J.E.,
 Lapiere L., Mollet B.;
 "A new cell surface proteinase: sequencing and analysis of the prtb
 gene from Lactobacillus delbrueckii subsp. bulgaricus";
 J. Bacteriol. 178:3053-3065(1996).
 R EMBL; L48487; AAC41529.1;
 R HSSP; P00782; ISUE.
 R Interpro; IPR00209;
 R Interpro; IPR003137;
 R Pfam; PF02225; PA; 1.
 R Pfam; PF02225; Peptidase_S8; 3.
 R PRINTS; PR00723; SUBTILISIN.
 R PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
 R PROSITE; PS00138; SUBTILASE_SER; 1.
 SIGNAL.

CHAIN 193 1946 POTENTIAL.
 PROTEINASE.
 SEQUENCE 1946 AA; 212315 MW; 21EF17D02E79C6A0 CRC64;

Query Match 11.5%; Score 82; DB 2; Length 1946;
 Best Local Similarity 23.5%; Pred. No. 11e+02;
 Matches 32; Conservative 17; Mismatches 31; Indels 56; Gaps 5;

28 VDSQSGSKSENLPP-----DQVSQKEGIAQEQ-----IVIKITDQGYTSHG 70
 171 IDVKVLPQVKNTLPKVVHPNDESADQMAQVQWQBGKAKGSAVISTIDIGDSSHQ 230
 71 D-----HYHYNGKVPYDALFSEELMKDPYOLKADIYV 106
 231 DLKLDGVSATLSKSEVSDSKSLGKGYTEKVPYGY-----NYDKNGQIYD 279
 107 -----EVKGGIIVDQ 118

DB 280 NGCGEMHGOHVAGING 295

RESULT 9

016944 PRELIMINARY; PRT; 703 AA.
 AC 016944;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE PROHORMONE CONVERTASE 1.
 OS Aplysia californica (California sea hare).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;
 OC Aplysiidae; Aplysia.
 X NCBI_Taxid=6500;

SEQUENCE FROM N.A.
 MEDLINE-96213914; PubMed-8639270;
 RA Gorham E.L., Nagle G.T., Smith J.S., Shen H., Kurosky A.;
 "Molecular cloning of prohormone convertase 1 from the atrial gland of
 Aplysia";
 RT DNA Cell Biol. 15:339-345(1996).
 DR EMBL; U04081; AAB06591.1; .
 DR HSSP; Q99405; IMPT.
 DR Interpro; IPR00209;
 DR Interpro; IPR00504;
 DR Interpro; IPR002884;
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF01483; P; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PRODOM; PD000717; .
 DR PROSITE; PS00030; RNP_1; UNKNOWN_1.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR SEQUENCE 703 AA; 77575 MW; 208C9D7006088A38 CRC64;

Query Match 11.5%; Score 81.5; DB 5; Length 703;
 Best Local Similarity 20.9%; Pred. No. 35;
 Matches 31; Conservative 21; Mismatches 39; Indels 57; Gaps 4;

14 HROQENKNNRVYVDGSSQSKSE----- 38
 DB 86 HHRKLSSEDRVAFVQDQKKRYKGLVDELDRELHRELARETAAGGELHDELHENT 145
 QY 39 -NLPQVSG-----KEGIAQEQYIVIKITDQGYTSHGDHYHYNGKVPYD 83
 DB 146 LNPQGEVSRSDVGRADLGKAVMKKGTGKIVTILLDDGIERTHPDLKSNYDPASVD 205
 QY 84 ALFSEELMKDPYOLKADIYNEVKG 111
 DB 206 --FNDEDEDPSPRY-----DITNENKRG 226

RESULT 10

016971 PRELIMINARY; PRT; 712 AA.

AC 016971;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE PCIB.
 GN Aplysia californica (California sea hare).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;
 OC Aplysiidae; Aplysia.
 X NCBI_Taxid=6500;
 RP SEQUENCE FROM N.A.
 MEDLINE-94213751; PubMed-8161455;
 RX Chun J.Y., Korner J., Kreiner T., Scheller R.H., Axel R.;

RT The function and differential sorting of a family of aplysia
 RT prothormone processing enzymes.
 RL Neuron 12:831-844(1994).
 DR EMBL: L28767; AAA27768.1;
 DR HSP: 099405; IMF.
 DR InterPro: IPR000209;
 DR InterPro: IPR000504;
 DR InterPro: IPR002884;
 DR Pfam: PF00082; peptidase_s8; 1.
 DR Pfam: PF01483; P; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PRODOM: PD000717; -; 1.
 DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 DR PROSITE: PS00136; SUBTILASE_CASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 SC SEQUENCE 712 AA; 78526 MW; C64957F3C61E860F CRC64;

Query Match 11.4%; Score 81; DB 5; Length 712;
 Best Local Similarity 20.8%; Pred. No. 39;
 Matches 31; Conservative 21; Mismatches 39; Indels 58; Gaps 4;

QY 14 HRSQENKDNKRVSYVDGSSQSKSE----- 38
 DB 94 HNRKLSSEERVAFEVQOQKRRKGLVEDRELHRELARETAAAGGELHDELHHEM 153
 QY 39 --NLTPDOVSQ-----REGIOAEQIVIKITDOGVTSHGHHYHNGKRVY 82
 DB 154 YLNPITSESRSEVRADGVAVRWKKGITGKIVTITDGGIERHHPDLKSNYDEAST 213
 QY 83 DALFSEELMKDPNYQLKADIVNEVKG 111
 DB 214 D--FNDEDEPSRY-----DITENKMG 235

RESULT 11
 QY 092M97 PRELIMINARY; PRT; 567 AA.
 AC 092M97;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 GN FLAGELLAR M-RING PROTEIN.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.S.L., Molt D.T., Brown E.D., Doly P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummiano P.J., Caruso A., Ulla-Nickelsen M., Mills D.M., Yves C.,
 RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Voyts G.F.,
 RA Trust T.J.;
 RT Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.
 RL Nature 397:176-180(1999).
 RU EMBL: AE001468; AAD05899.1;
 DR EMBL: AE001468; AAD05899.1;
 DR InterPro: IPR000067;
 DR InterPro: IPR002920;
 DR Pfam: PF01514; YscJ; 1.
 DR PRINTS: PR01009; FLGMRINGFLIF.
 SC SEQUENCE 567 AA; 63563 MW; 458B99EF2C85CB24 CRC64;

Query Match 11.2%; Score 79.5; DB 2; Length 567;
 Best Local Similarity 23.9%; Pred. No. 40;
 Matches 38; Conservative 22; Mismatches 52; Indels 47; Gaps 7;

QY 10 ALMOHNSQENKDN-----NRV-----SYVDSQSSQSKSENLTPDY-- 45

DB 225 ALEQLRKQNFENLEKNTLVNITLAPVGGKKNVAVRVNAEFDPSQKSTKEFDENNVVR 294
 QY 46 -----SQREGIOAEQIVIKITDOGVTSHGHHYHNGKRVYDALLFSEELMKDPNYQL 99
 DB 295 SQONLEEKKEGAPKQVQV--GVPGVSNIGFVQGLKDKKEE---KYKSQNTTNEV 347
 QY 100 KADIVNEVKGXY-----IIRYDKKRYVYLKDAHA 130
 DB 348 --GKTISELKEGEGTLVRLNAVVDGKIKKILDEGANA 384

RESULT 12
 QY 096185 PRELIMINARY; PRT; 2573 AA.
 AC 096185;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE HYPOTHETICAL 308.1 KDA PROTEIN.
 GN PF00460C.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99021743; PubMed=9804551;
 RA Gardner M.J., Tetteilin H., Carucci D.J., Cummings L.M., Aravind L.,
 RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
 RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perlea M.,
 RA Salberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
 RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
 RT Chromosome 2 sequence of the human malaria parasite Plasmodium
 RT falciparum.
 RL Science 282:1126-1132(1998).
 DR EMBL: AE001396; AAC71881.1;
 KW Hypothetical protein.
 SC SEQUENCE 2573 AA; 308142 MW; 49E0D7E8967066CF CRC64;

Query Match 11.2%; Score 79.5; DB 5; Length 2573;
 Best Local Similarity 20.8%; Pred. No. 2,5e+02;
 Matches 31; Conservative 25; Mismatches 38; Indels 55; Gaps 6;

QY 12 NQRSQENKDNKRVSYVDGSSQSKSE-----NLTPD-----QY 45
 DB 2421 NQNN 2480
 QY 46 SQREGIOAEQIVIKITDOGVTSHG---DHYH-----YNGKVPYDA 84
 DB 2481 TCKESYEVEVEIITKTKRKF--HNIELEKHYCYDLFKRKLENTYRNTYKRRKILINN 2537
 QY 85 LFESEELMKDPNYQLKADIVNEVKGXY 113
 DB 2538 LITNKNI-----FYKEHDIYVNVKQIF 2561

RESULT 13
 QY 095937 PRELIMINARY; PRT; 264 AA.
 AC 095937;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE LIPASE-ESTERASE (EC 3.1.1.3) (TRIGLYCEROL LIPASE) (TRIGLYCERIDE
 DE LIPASE) (TRIBUTYRASE).
 GN Lip1.
 OS Mycoplasma mycoides.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Entomoplasmataceae.
 OX NCBI_TaxID=2102;
 RN [1]
 RP SEQUENCE FROM N.A.

[illegible]

```

Query Match          11.0%; Score 78; DB 2; Length 438;
Best Local Similarity 28.6%; Pred. No. 40;
Matches      26; Conservative      10; Mismatches      27; Indels      28; Gaps      4

QY      58  IRTDGYTSHGDHYHYNGKRPYDAL---ESELAMD-----PNYQKDA- 102
      | : : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      278  INVQCPGMMTQNGSKTLGDEKRPYDILSLFGDDLELSAKKVGSPRSYFGYTLKTGY 337
      | : : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      103  -----DIVNEYKGGY-----IIRVDGKY 120
      : : : | : | : | : | : | : | : | : | : | : | : | : |
DB      338  SGEPRVIVIQGLNMAISRAVPLPKTIAVDGKY 368

```

RESULT	ID	QPPHM6	PRELIMINARY;	PRT:	642 AA.
09PHW6	ID	QPPHM6	PRELIMINARY;	PRT:	642 AA.
AC	09PHW6	01-OCT-2000 (TREMBLrel, 15, Created)			
DT	01-OCT-2000 (TREMBLrel, 15, last sequence update)				
DT	01-MAR-2001 (TREMBLrel, 16, last annotation update)				
DE	POTATIVE FLAGELLAR HOOK-ASSOCIATED PROTEIN.				
GN	FLID OR CJ0548.				
OS	Campylobacter jejuni.				
CC	Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group; Campylobacter.				
OX	NCBI_taxid=197;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=NCIC 11168;				
FX	MEDLINE=20150912; PubMed=10686204;				
RA	Parkhill J., Wen B.W., Mungall K., Ketley J.M., Churcher C.,				
RA	Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,				
RA	Jegels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,				
RA	Quail M.A., Rajandream M.A., Rutherford K.M., Van Vleet A.H.M.,				
RA	Whitehead S., Barrall B.G.;				
RT	"The genome sequence of the food-borne pathogen Campylobacter jejuni				
RT	reveals hypervariable sequences."				
RL	Nature 403:665-668(2000)				
DR	EMBL: AL139075; CMB5184.1;				
DR	InterPro: IPR003481;				
DR	Pfam: PF02465; FLID; 1.				
SO	SEQUENCE	642 AA; 69775 MW; 40FE20DA1668EFA3 CRC64;			

	Query Match	11.0%	Score 78;	DB 2;	Length 642;
	Best Local Similarity	28.3%	Pred. No. 63;		
	Matches 28;	Conservative 16;	Mismatches 39;	Indels 16;	Gaps 4;
QY	21 DNNRSTYVDGSSQSSKNLPPDYOSKEGIOAEIVIKITGGVYTSH---- <td> </td> <td> </td> <td> </td> <td> </td>				
DB	84 DNPASLTVNSGVALQSMKNINYTOAQDYVOSKILA---NNGSGVNAOLNGTADLTFFS 140				
QY	77 NGKVPYDALFSEELLMKDPNTQLRK-AIVIVEKVGYYII 114				
DB	141 NGK-----EYVTVYDKNTTYTRDLAKINKINASGGEIV 171				

Search completed: September 26, 2001, 22:33:48
Job time: 267 sec

ALL INFORMATION CONTAINED
HEREIN IS UNCLASSIFIED
DATE 09-11-2001 BY 60322
UCBAW

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2001, 22:31:17 ; Search time 32.54 seconds
(without alignments)
86.057 Million cell updates/sec

Title: us-09-471-255-2_copy_15_150

Perfect score: 711

Sequence: 1 IVSLICAYALNHRQENK.....DGKYYVYLKDAHADNVRTK 136

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents AA:*

- 1: /cgn2.6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2.6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2.6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2.6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2.6/ptodata/2/1aa/PTUS_COMB.pep.*
- 6: /cgn2.6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Match	Length	ID	Description
1	663	93.2	447	US-08-961-083-182	Sequence 182, App
2	575.5	80.9	763	US-08-961-083-66	Sequence 66, Appl
3	574	80.7	796	US-08-961-083-56	Sequence 56, Appl
4	71.5	10.1	289	US-08-284-941-10	Sequence 10, Appl
5	71.5	10.1	289	US-08-447-642-10	Sequence 10, Appl
6	71.5	10.1	289	PCT-US93-02147A-10	Sequence 10, Appl
7	71.5	10.1	753	US-08-712-241-3	Sequence 3, Appl1
8	71.5	10.1	753	PCT-US92-10621-3	Sequence 3, Appl1
9	71.5	10.1	753	PCT-US94-02233-3	Sequence 3, Appl1
10	69.5	9.8	208	US-08-896-933-30	Sequence 30, Appl
11	69.5	9.8	208	US-08-480-604A-21	Sequence 21, Appl
12	69.5	9.8	608	US-08-405-496A-21	Sequence 21, Appl
13	69.5	9.8	609	US-08-480-604A-10	Sequence 10, Appl
14	69.5	9.8	2366	US-08-480-604A-10	Sequence 10, Appl
15	69.5	9.8	2366	US-08-405-496A-10	Sequence 10, Appl
16	68.5	9.6	753	US-08-712-241-2	Sequence 2, Appl1
17	68	9.6	178	US-08-650-528-2	Sequence 2, Appl1
18	68	9.6	178	US-09-060-584-2	Sequence 2, Appl1
19	68	9.6	178	US-09-413-140A-2	Sequence 2, Appl1
20	67.5	9.5	519	US-08-997-445D-2	Sequence 2, Appl1
21	66.5	9.4	203	US-08-801-740-8	Sequence 8, Appl1
22	66.5	9.4	203	US-08-801-740-8	Sequence 8, Appl1
23	66	9.3	248	US-09-067-800-6	Sequence 6, Appl1
24	65.5	9.2	520	US-08-961-083-122	Sequence 122, App
25	65.5	9.2	775	US-07-603-133B-14	Sequence 14, Appl
26	65.5	9.2	923	US-08-936-135-6	Sequence 6, Appl1
27	65.5	9.2	1040	US-08-961-083-118	Sequence 118, App

Issued Patents

28	65.5	9.2	1528	1	US-08-326-117B-2	Sequence 2, Appl1
29	65.5	9.2	1528	3	US-08-982-129-2	Sequence 2, Appl1
30	65	9.1	378	4	US-08-975-762-70	Sequence 70, Appl
31	64.5	9.1	382	2	US-08-504-265B-75	Sequence 75, Appl
32	64.5	9.1	798	1	US-08-190-802A-64	Sequence 64, Appl
33	64.5	9.1	798	1	US-08-190-802A-68	Sequence 68, Appl
34	64.5	9.1	798	2	US-08-308-818-2	Sequence 2, Appl1
35	64.5	9.1	798	2	US-07-728-215-30	Sequence 30, Appl
36	64	9.0	462	1	US-08-267-092A-1	Sequence 1, Appl1
37	64	9.0	462	2	US-08-540-412-1	Sequence 1, Appl1
38	64	9.0	462	4	US-09-051-342-1	Sequence 1, Appl1
39	64	9.0	462	4	US-08-468-161-1	Sequence 1, Appl1
40	64	9.0	462	4	US-09-051-759-1	Sequence 1, Appl1
41	64	9.0	462	5	PCT-US95-08156-1	Sequence 1, Appl1
42	64	9.0	522	6	RE34608-6	Patent No. RE34,60
43	64	9.0	560	2	US-09-132-619-10	Sequence 10, Appl
44	64	9.0	560	3	US-09-282-803B-10	Sequence 10, Appl
45	64	9.0	1112	2	US-08-714-402-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-961-083-182
; Sequence 182, Application US/08961083
; Patent No. 6159463
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-182

Query Match 93.2%; Score 663; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.8e-66;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 LNHRSQENKNNRVSYDGSOSKSENLPDQVSKESIQAEQIVKITDQGYVTSHG 70
|||||

Db 1 LNHRSQENKNNRVSYDGSQSSQKSENTPDOVSQKEGIAEQIVIKITDGYTSHG 60
QY 71 DHHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGIIIRKDYKYYVLKDAHA 130
Db 61 DHHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGIIIRKDYKYYVLKDAHA 120
QY 131 DNVRTK 136
Db 121 DNVRTK 126

RESULT 2

US-08-961-083-66
Sequence 66, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Chol et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-66

Query Match 80.9%, Score 575.5, DB 4, Length 763,
Best Local Similarity 78.6%, Pred. No. 2.7e-56,
Matches 103; Conservative 18; Mismatches 9; Indels 1; Gaps 1;

QY 7 CAYALNQRS-OENKNNRVSYDGSQSSQKSENTPDOVSQKEGIAEQIVIKITDGYT 65
Db 1 CSEYELGRHQAGVKNRSNRSYIDGDAQKRAENLTPDEVSKREGINAEQIVIKITDGY 60
QY 66 VHSQDHHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGIIIRKDYKYYVLK 125
Db 61 VHSQDHHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGIIIRKDYKYYVLK 120
QY 126 DAHADNVRTK 136
Db 121 DAHADNVRTK 131

RESULT 3

US-08-961-083-56
Sequence 56, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Chol et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-56

Query Match 80.7%, Score 574, DB 4, Length 796,
Best Local Similarity 80.6%, Pred. No. 4.3e-56,
Matches 104; Conservative 16; Mismatches 9; Indels 0; Gaps 0;

QY 8 AYALNQRSQENKNNRVSYDGSQSSQKSENTPDOVSQKEGIAEQIVIKITDGYT 67
Db 1 SYELGLYQARFYKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDGYT 60
QY 68 SHQDHHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGIIIRKDYKYYVLKDA 127
Db 61 SHQDHHYHNGKVPYDALFSEELMKDPNYQLKADIVNEVKGIIIRKDYKYYVLKDA 120
QY 128 AHADNVRTK 136
Db 121 AHADNVRTK 129

RESULT 4
US-08-284-941-10
Sequence 10, Application US/08284941
Patent No. 5863756

GENERAL INFORMATION:
APPLICANT: BARR, PHILIP J
APPLICANT: KIEFER, MICHAEL C
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODWARD CASTRO, HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE

```

Query Match      10.1%  Score 71.5:  DB 2,  Length 289;
Best Local Similarity 31.2%:  Pred. No. 3.2;
Matches 20;  Conservative 11;  Mismatches 26;  Indels 7;  Gaps 21;

OY 48 KEGIOAEIVIKITDQGVVTSHGDIHYHYNGKRVDPALFSEELMKKDPYQGLKADIYNE 107
Db 32 EKGITGKGVAVTVLDDGLENNHTDIYANYDPEAST--FNDNDHPDPFY-----DLTNE 84

OY 108 VKGG 111
Db 85 NKHG 88 /

```

NAME: NEELEY PH. D., RICHARD L.
REGISTRATION NUMBER: 30092
REFERENCE/DOCKET NUMBER: CHIR-009/01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 843-5070
TELEFAX: (415) 857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-447-642-10

```

Query Match      10.1%; Score 71.5; .DB 2; Length 289;
Best Local Similarity 31.2%; Pred. No. 3.2;
Matches 20; Conservative 11; Mismatches 7; Gaps 2;

OY 48 KEGIOAEIVIKITDQGVTSHGDPHYHYNGKVPYDALFSEELMKDNPYOLKADIVNE 107
Db 32 EKGITGKGVAVITVDDGLENNHTDIYANIPEASTD--FNDNDHPFPFY-----DLTNE 84

OY 108 VKGG 111
Db 85 NKHG 88

```

RESULT 6
 PCT-US93-02147A-10
 Sequence 10: Application PC/TUS9302147A
 GENERAL INFORMATION:
 APPLICANT: BARR, PHILIP J
 APPLICANT: KIEFER, MICHAEL C
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PAGE 4 AND
 TITLE OF INVENTION: PAGE 4.1 GENE AND POLYPEPTIDES IN CELLS
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
 STREET: FIVE PALO ALTO SQUARE
 CITY: PALO ALTO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/02147A
 FILING DATE: 19930309
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,629
 FILING DATE: 09-MAR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: NEEDLEY PH.D., RICHARD L
 REGISTRATION NUMBER: 30092
 REFERENCE/DOCKET NUMBER: CHR-009/00US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 494-7622
 TELEFAX: (415) 857-0663
 TELEX: 380816 COOLEY PA
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 289 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US93-02147A-10

Query Match 10.1%; Score 71.5; DB 5; Length 289;
Best Local Similarity 31.2%; Pred. No. 3.2;
Matches 20; Conservative 11; Mismatches 26; Indels 7; Gaps 2;

OY 48 KEGIOAEQIVIKITDGYTSHGDHYHYNGKVPYALFSEELMDPNYQLADADIVNE 107
DB 32 EKGITGKGVIVYLDGLENHTDIYANDPEASID--FNDNDHDPFPRY-----DLTNE 84

OY 108 VKGG 111
DB 85 NKG 88

RESULT 7
US-08-712-241-3
Sequence 3, Application US/08712241
Patent No. 5789564

GENERAL INFORMATION:

APPLICANT: SEIDAH, NABIL G.

TITLE OF INVENTION: DEVELOPMENT OF RESEARCH,

TITLE OF INVENTION: DIAGNOSTIC AND PRODUCTION TOOLS FOR PRO-HORMONE

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSER: QUARLES & BRADY

STREET: 411 EAST WISCONSIN AVENUE

CITY: MILWAUKEE

STATE: WISCONSIN

COUNTRY: U.S.A.

ZIP: 53202-4497

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 Inch, 720 kb diskette

COMPUTER: IBM PS/2, Model 30

OPERATING SYSTEM: PC-DOS 3.30

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/712,241

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/529,785

FILING DATE: OCTOBER 20, 1992

APPLICATION NUMBER: US/07/963,535A

FILING DATE: OCTOBER 20, 1992

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: JEAN C. BAKER

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 20-702-9001-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (414) 277-5000

TELEFAX: (414) 277-5591

TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 753 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-712-241-3

Query Match 10.1%; Score 71.5; DB 1; Length 753;

Best Local Similarity 31.2%; Pred. No. 13;

Matches 20; Conservative 11; Mismatches 26; Indels 7; Gaps 2;

OY 48 KEGIOAEQIVIKITDGYTSHGDHYHYNGKVPYALFSEELMDPNYQLADADIVNE 107
DB 153 EKGITGKGVIVYLDGLENHTDIYANDPEASID--FNDNDHDPFPRY-----DLTNE 205
OY 108 VKGG 111

DB 206 NKG 209

RESULT 8
PCT-US92-10621-3
Sequence 3, Application PC/TUS9210621

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: Gorman, Cornelia M.

APPLICANT: Mariotti, Dave

APPLICANT: Groskreutz, Debra J.

TITLE OF INVENTION: Prohormone Convertase Transformed Cells and Polypeptide Sy

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSER: Genentech, Inc.

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 Inch, 360 kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Datal (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/10621

FILING DATE: 19921204

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/887265

FILING DATE: 22-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/803631

FILING DATE: 06-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Adler, Carolyn R.

REGISTRATION NUMBER: 32,324

REFERENCE/DOCKET NUMBER: 748P2, PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-2614

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 753 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

PCT-US92-10621-3

Query Match 10.1%; Score 71.5; DB 5; Length 753;

Best Local Similarity 31.2%; Pred. No. 13;

Matches 20; Conservative 11; Mismatches 26; Indels 7; Gaps 2;

OY 48 KEGIOAEQIVIKITDGYTSHGDHYHYNGKVPYALFSEELMDPNYQLADADIVNE 107
DB 153 EKGITGKGVIVYLDGLENHTDIYANDPEASID--FNDNDHDPFPRY-----DLTNE 205
OY 108 VKGG 111
DB 206 NKG 209

RESULT 9
PCT-US94-02233-3
Sequence 3, Application PC/TUS9402233

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

TITLE OF INVENTION: Prohormone Convertase Transformed Cells and Polypeptide Sy

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/02233
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 748P3PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 753 amino acids
TYPE: amino acid
TOPOLOGY: linear
PCT-US94-02233-3

Query Match 10.1%; Score 71.5; DB 5; Length 753;
Best Local Similarity 31.2%; Pred. No. 13;
Matches 20; Conservative 11; Mismatches 26; Indels 7; Gaps 2;

QY 48 KEGIOABQIVIKITDGYVSHGDHYHYNGKVPYDALFSEELMKDPNTQLKDADIVNE 107
DB 153 KEGITGKGVYTVLDDLENNHTDIYANYDPEASVD--FNDNDHDPPEPRY-----DLTNE 205
108 VKGG 111

DB 206 NKGK 209

RESULT 10
US-08-896-933-30
Sequence 30, Application US/08896933
Patent No. 6221351
GENERAL INFORMATION:
APPLICANT: Terman, David S.
TITLE OF INVENTION: TUMOR KILLING EFFECTS OF ENTEROTOXINS,
FILE REFERENCE: 09629/005002
CURRENT APPLICATION NUMBER: US/08/896,933
CURRENT FILING DATE: 1997-07-18
EARLIER APPLICATION NUMBER: 08/252,978
EARLIER FILING DATE: 1994-06-02
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 30
LENGTH: 208
TYPE: PRT
ORGANISM: Streptococcus pyogenes
US-08-896-933-30

Query Match 9.8%; Score 69.5; DB 4; Length 208;
Best Local Similarity 22.6%; Pred. No. 3.3;
Matches 33; Conservative 25; Mismatches 47; Indels 41; Gaps 8;

QY 9 YALNDR-----SOENKDNRY---STYDSSOSSCKSEBNTPPDOVSQREGIOABQ 55
DB 76 YILNSHTGETYIGITPAQNNKYNHKLGNLFISG-----ESQNNLNKLTLELDYTFQE 131
QY 56 IYIKITDGYVSHGDHYHYNGKVPY-----DALFSEELMKDPNTQLKDADIV 105
DB 132 IDEFKI--RYLW---DNRYITDTSPTVSGRLEIGTKDKNHEQIDLFDSPNEGTR--SDIF 185
QY 106 NEYKGGYIIRKDGKRYVYLKDAHAD 131
DB 186 AKTKDNRIIN-----MKNSHFD 203

RESULT 11
US-08-480-604A-21
Sequence 21, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLTA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 608 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown

MOLECULE TYPE: protein
US-08-480-604A-21

Query Match 9.8%; Score 69.5; DB 1; Length 608;
Best Local Similarity 22.1%; Pred. No. 16;
Matches 27; Conservative 26; Mismatches 42; Indels 27; Gaps 6;

QY 16 SOENKNN-RVSYV-----GOSOSKSENLPDQVSKGIGIAEQIVIKT-----61
DB 1 SEENKVSQKIRVNFVKDITLANKLSFN-----SDKQDVPVSEIILSFTPSYEDGLI 55
QY 62 --DGGVTSHGDIHYHNGKVPDADFSEELLKDPNOLKADIVNEVKGIIIVDGR 119
DB 56 GYDLGLVSLYNEKFYINN-----FGMAYSGLIYINDSLYFKRP--VNNLTGVTYVGDDK 109
QY 120 YY 121
DB 110 YY 111

RESULT 12

US-08-405-496A-21
Sequence 21, Application US/08405496A
Patent No. 5919665

GENERAL INFORMATION:

APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEOROTOXIN
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 608 amino acids
TYPE: amino acid
STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-405-496A-21

Query Match 9.8%; Score 69.5; DB 2; Length 608;
Best Local Similarity 22.1%; Pred. No. 16;
Matches 27; Conservative 26; Mismatches 42; Indels 27; Gaps 6;

QY 16 SOENKNN-RVSYV-----GOSOSKSENLPDQVSKGIGIAEQIVIKT-----61
DB 1 SEENKVSQKIRVNFVKDITLANKLSFN-----SDKQDVPVSEIILSFTPSYEDGLI 55
QY 62 --DGGVTSHGDIHYHNGKVPDADFSEELLKDPNOLKADIVNEVKGIIIVDGR 119
DB 56 GYDLGLVSLYNEKFYINN-----FGMAYSGLIYINDSLYFKRP--VNNLTGVTYVGDDK 109
QY 120 YY 121
DB 110 YY 111

RESULT 13

US-08-480-604A-30
Sequence 30, Application US/08480604A
Patent No. 5736139

GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER:

REFERENCE/DOCKET NUMBER: OPD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338

REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPHD-01308
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2366 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-403-496A-10

Query Match 9.8%; Score 69.5; DB 2; Length 2366;
 Best Local Similarity 22.1%; Pred. No. 1.2e+02;

Matches 27; Conservative 26; Mismatches 42; Indels 27; Gaps 6;

QY 16 SOEKDNN-RVSYV---DGSOSQKSENTLPQVSQKREGIOAEQIVIKIT-----61

Db 1755 SEEKVSQVIRRVNPFKDKTLANKLSFN-----SDKQVPVSEITLSTPSIYEDGLI 1809

QY 62 --DQGYTSGDHYHYNGRVYPDALFSELLKMDPNYOLKADIVNEVNGGYIIRVDGR 119

Db 1810 GYDLGLVSLYNEFYINN---FGMVSGLIYINDSLYFKPP--VNNLITGVTYGDGR 1863

120 YR 121

Db 1864 YR 1865

Search completed: September 26, 2001, 22:31:19
 Job time: 1543 sec

Thu Sep 27 09:15:58 2001

us-09-471-255-2.ra1

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: September 26, 2001, 22:05:36 ; Search time 32.54 Seconds

(without alignments)
657,448 Million.cell updates/sec

Title: US-09-471-255-2

Perfect score: 5406

1 MFPSKRYTLAGSIVYSL.....IEKRLPSGEVYKNSLDFIA (2039)

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 197339 segs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCPUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfill1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	2338	43.2	447	US-08-961-083-182	Sequence 182, App
2	1247	23.1	796	US-08-961-083-56	Sequence 56, Appl
3	1228.5	22.7	763	US-08-961-083-66	Sequence 66, Appl
4	223	4.1	1964	US-08-790-912-3	Sequence 3, Appl
5	223	4.1	2052	US-08-790-912-2	Sequence 2, Appl
6	205.5	3.8	1183	US-08-447-031A-2	Sequence 2, Appl
7	203	3.8	1848	US-08-296-791-6	Sequence 6, Appl
8	203	3.8	1848	PCT-US95-10661A-6	Sequence 6, Appl
9	200	3.7	571	US-08-961-083-4	Sequence 4, Appl
10	191	3.5	1507	5268270-2	Patent No. 5268270
11	186	3.4	1231	US-08-904-263A-4	Sequence 4, Appl
12	178.5	3.3	2465	US-08-596-291-3	Sequence 3, Appl
13	178.5	3.3	2465	US-09-100-804-3	Sequence 3, Appl
14	177	3.3	1702	US-08-296-791-5	Sequence 5, Appl
15	177	3.3	1702	PCT-US95-10661A-5	Sequence 5, Appl
16	175.5	3.2	2466	US-09-080-855-12	Sequence 12, Appl
17	175.5	3.2	2466	PCT-US94-09943-2	Sequence 2, Appl
18	174.5	3.2	2485	US-09-290-640-46	Sequence 46, Appl
19	173.5	3.2	1780	US-08-769-309A-5	Sequence 5, Appl
20	173.5	3.2	1780	US-08-994-570-5	Sequence 5, Appl
21	171	3.2	984	US-08-242-932-2	Sequence 2, Appl
22	171	3.2	984	US-08-714-481-2	Sequence 2, Appl
23	171	3.2	984	PCT-US95-06111-2	Sequence 2, Appl
24	170.5	3.2	1043	US-08-928-361B-30	Sequence 30, Appl
25	170.5	3.2	1721	US-08-700-651-5	Sequence 5, Appl
26	170.5	3.2	1721	US-08-928-361B-6	Sequence 6, Appl
27	169.5	3.1	1167	US-08-589-756-2	Sequence 2, Appl

28	168.5	3.1	1435	US-08-568-459A-4	Sequence 4, Appl
29	168.5	3.1	1435	US-08-487-826B-4	Sequence 4, Appl
30	168.5	3.1	1566	US-08-687-956A-23	Sequence 23, Appl
31	165.5	3.1	1007	US-08-961-083-216	Sequence 216, App
32	163.5	3.0	1545	US-08-296-791-4	Sequence 4, Appl
33	163.5	3.0	1545	PCT-US95-10661A-4	Sequence 4, Appl
34	162.5	3.0	783	5231168-2	Patent No. 5231168
35	162.5	3.0	1164	US-08-589-756-1	Sequence 1, Appl
36	162	3.0	2409	5180808-2	Patent No. 5180808
37	161.5	3.0	1529	US-08-728-470-10	Sequence 10, Appl
38	161.5	3.0	1529	US-08-719-641-10	Sequence 10, Appl
39	160.5	3.0	1600	US-08-617-697-10	Sequence 10, Appl
40	159.5	3.0	1085	US-08-431-080-28	Sequence 28, Appl
41	159.5	3.0	1085	US-08-938-534-28	Sequence 28, Appl
42	159	2.9	1042	US-08-928-361B-11	Sequence 11, Appl
43	159	2.9	1837	US-08-928-361B-5	Sequence 5, Appl
44	158.5	2.9	802	US-09-156-316-1	Sequence 1, Appl
45	155.5	2.9	2353	US-09-377-155-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-08-961-083-182
Sequence 182, Application US/08961082
Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Chol et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS: ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM: MEDIUM TYPE: Diskette, 3.50 Inch, 1.4MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA: APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA: APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION: NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION: TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 182:

SEQUENCE CHARACTERISTICS:

LENGTH: 447 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-961-083-182

Query Match 43.2% Score 2338, DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 2e-156;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
25 INORSDENKNNRVSYSYSSSSSSKSENLPPDOVSOREGIGAOIVKIRIDGQVTSHG 84
|||||

Db 1 LNORSEKNDNNVSYVDSQSSQSEKNTLPDQVSOKEGIAQOYIKITDQITVSHG 60
 QY 85 DRYHYNGKPYDALFSEELMKDPNTOLKADIVNEVKGKGIIRKVDGKYYVYLKDAHA 144
 Db 61 DRYHYNGKPYDALFSEELMKDPNTOLKADIVNEVKGKGIIRKVDGKYYVYLKDAHA 120
 QY 145 DNVTKEINROKOEHYDKNEKVNNAVAASQGRYTTNDGYVNPADIIEDTGNATV 204
 Db 121 DNVTKEINROKOEHYDKNEKVNNAVAASQGRYTTNDGYVNPADIIEDTGNATV 180
 QY 205 HGHYHTYPKDLSASELAAKALAKKNNOPSOLOSTSTASDNNWTSVAKGSKPANK 264
 Db 181 HGHYHTYPKDLSASELAAKALAKKNNOPSOLOSTSTASDNNWTSVAKGSKPANK 240
 QY 265 SENOSLKELYDSPAQRYSSESDGLVDPKAITSRTPNGVAPHGHDHYHTPYKSLAL 324
 Db 241 SENOSLKELYDSPAQRYSSESDGLVDPKAITSRTPNGVAPHGHDHYHTPYKSLAL 300
 QY 325 EKIARWVPISGTSVSTNAKPNEVYSSLSLSSNPSSILTSKELSSASDGYTFNPDI 384
 Db 301 EKIARWVPISGTSVSTNAKPNEVYSSLSLSSNPSSILTSKELSSASDGYTFNPDI 360
 QY 385 VEETATATVHGHFHFIPIPSNIOGPTLPNNSLATPSPLPINPGTSHKHEHDGYG 444
 Db 361 VEETATATVHGHFHFIPIPSNIOGPTLPNNSLATPSPLPINPGTSHKHEHDGYG 420
 QY 445 DANRIADESGFVMSHGDHNYEPK 471
 Db 421 DANRIADESGFVMSHGDHNYEPK 447

RESULT 2

US-08-961-083-56

Sequence 56, Application US/08961083

Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Chol et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:

LENGTH: 796 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-961-083-56

Query Match 23.18; Score 1247; DB 4; Length 796;
 Best Local Similarity 41.28; Pred. No. 1,9e-79;
 Matches 298; Conservative 95; Mismatches 180; Indels 150; Gaps 19;

QY 22 AYALNORSEKNDNNVSYVDSQSSQSEKNTLPDQVSOKEGIAQOYIKITDQITV 81
 Db 1 SYELGLYQARVYKNNVSYVDSQSSQSEKNTLPDQVSOKEGIAQOYIKITDQITV 60
 QY 82 SHGDHYHYNGKPYDALFSEELMKDPNTOLKADIVNEVKGKGIIRKVDGKYYVYLKDA 141
 Db 61 SHGDHYHYNGKPYDALFSEELMKDPNTOLKADIVNEVKGKGIIRKVDGKYYVYLKDA 120
 QY 142 AHADNVTKKEINROKOEHYDKNEKVNNAVAASQGRYTTNDGYVNPADIIEDT 197
 Db 121 AHADNVTKKEINROKOEHYDKNEKVNNAVAASQGRYTTNDGYVNPADIIEDT 180
 QY 198 GNATVYHGHFHFIPIPSNIOGPTLPNNSLATPSPLPINPGTSHKHEHDGYG 248
 Db 181 GNATVYHGHFHFIPIPSNIOGPTLPNNSLATPSPLPINPGTSHKHEHDGYG 240
 QY 249 -----NTOSVAKGSTRPANKSENOSLKELYDSPAQRYSSESDGLVDPKAITSR 301
 Db 241 VSNPCTNTNTSNNSTNSQASNDISLKOYLKPLSGRHVESDGLVDPKAITSR 300
 QY 302 PNGVAIPHGHFHFIPIPSNIOGPTLPNNSLATPSPLPINPGTSHKHEHDGYG 333
 Db 301 ARGVAIPHGHFHFIPIPSNIOGPTLPNNSLATPSPLPINPGTSHKHEHDGYG 360
 QY 334 -----ISGTSVSTNAK-----PNEVYSSLSLSSNPSS 363
 Db 361 POPAPNKLIDNSSLVQVYKVEGVEEKGISRYFANDLSEYVKNLESKLSQES 420
 QY 364 -----LTSKELSSASDGYTFNPK-DIYEETATVYHGHFHFIPIPSNIOGPTLP 410
 Db 421 VSHVTLTAKKENVAPRDEFTYKAVNLTLEAHKALFYKNGRNSDQALDKLERLNDE 480
 QY 411 QPTLPNNSLATPSP-LP-INPGTSHKHE-----EDGYGFDAARRIAE 452
 Db 481 KEKLYDOLLAFLAPLITHERGKPNQSELEYEDERVIAQLADKYTSGYTFDEHDITSD 540
 QY 453 DESGFVMSHGDHNYEPKIDTEQOIKAAQNHLE-----YTSINGLDSLSH 501
 Db 541 EGDAYVTFPHMGSHWIGKDSLDREKVAQAQYTEKGIPLPSPADVAPANTG-D 599
 QY 502 EODYPGNKKEKNDKTEKIKIAGIMQYGRRESIYVNEKNAIYPHGDHNRADP 558
 Db 600 YNRVKG-----EKRIPLVRLPYV--EHTYVAKNGNLI-PRHDHYHNIKFAW 644
 QY 559 IDEH---RPVGIG---HSHSNYELFKPEEGVAKKGNKYTGSELNVAVLKNTFN 611
 Db 645 FDDHYKAPNGYLTLEDLPAITKYVEHDERPHSDG-----WGNASEHYVLKDKHSEDP 700
 QY 612 QNF 614
 Db 701 KNF 703

RESULT 3

US-08-961-083-66

Sequence 66, Application US/08961083

Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Chol et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 Inch, 1.44MB storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,083
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 66:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 763 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-961-083-66

Query Match 22.7%; Score 1228.5; DB 4; Length 763;
 Best Local Similarity 40.5%; Pred. No. 3,6e-78;
 Matches 296; Conservative 101; Mismatches 185; Indels 149; Gaps 24;

21 CAYLALNQRS-OENKDNRRSYVDGSSQSKSENLPDQVSOKEGIAEQIVIKITPDQY 79
 1 CAYLALNQRS-OENKDNRRSYVDGSSQSKSENLPDQVSOKEGIAEQIVIKITPDQY 60
 80 VTSHGDIHYHYNGKVPYDALFSEELMKDPNOLKADIVNEVKGIIIVDGRYYLTK 139
 61 VTSHGDIHYHYNGKVPYDALFSEELMKDPNOLKADIVNEVKGIIIVDGRYYLTK 120
 140 DAHAADNVRKDEINROKOEHYVD-NKYNANVAVASOGRTYNTGNYVNPADITDTG 198
 121 DAHAADNVRKDEINROKOEHYVD-NKYNANVAVASOGRTYNTGNYVNPADITDTG 180
 199 NAYIVPHGHHYHYIPKSDLSASELAALAKAGKRNQPSOLYSTASDNNQ---SVAK 255
 181 DAYIVPHGHHYHYIPKSDLSASELAALAKAGKRNQPSOLYSTASDNNQ---SVAK 238
 256 GSTSKPA--NKSENLOSLKELYDPSAORYSESDGLVDPAKIISRTPNGVAIPRGDH 312
 239 NLVYTPYHONGENISILRELILYAKPLSRHVESDGLIFDPAQITSRARGVAVPHGNH 298
 313 YHFIPIYKLSALEKIRARWPIISGTGSTVSTNAKPN-----EYVSSLSLS 358
 299 YHFIPIYKLSALEKIRARWPIISGTGSTVSTNAKPN-----EYVSSLSLS 358
 359 SNPSLSLTKE--LSASADGIYFN-----PDIYEETAT--ATVHGHGHHF--- 402
 359 SNPSLSLTKE--LSASADGIYFN-----PDIYEETAT--ATVHGHGHHF--- 402
 403 ---TPKSN-----QIGQPTLPNNSLATPSPSL----- 426
 419 KRTDLPSSDEEFYKADLARIHODLLDNKRGVDFEALDNLLEKLDKDYSPKVLXVD 478
 427 ---TPKSN-----QIGQPTLPNNSLATPSPSL----- 426
 479 ILAVLAIIRPERLGRKNAOITTYDDIYQAKLAGKTTDEGYTFDPRDITSDGAYVT 538
 460 SHGDHNYFFPKDLTEQDIAQAQRHLEEVTSNGLDLSLSEHODYGNAKEMKMDLCK-- 517
 539 PHMTHSMIKKIDSLSEERAAAOVAKK-----KGLTPPSTDHQD-SGNT-EAKGALAIY 591

518 ---KIEEKIAGIMQYGVKRESIYVNEKNAIIPROGHHADPIDERKPVGISHSNTY 574
 592 NRYKAATVLDNRPNYLO---TYVEKNGSLIIPHIDHH-----NKF 633
 575 ELFPREGVAKKRCGNKTYTGELTNVNL-----KNSTENONFTL---ANGOKR 622
 634 EMF--DEGL--YEAPKGYTLEDLAIATVKKYVEHNERPHSDNGSNASDHQRRKNCQAD 689
 623 VSFSPPELEK 633
 690 TNQTERPSEER 700

RESULT 4
 US-08-790-912-3
 Sequence 3, Application US/08790912
 Patent No. 5976542
 GENERAL INFORMATION:
 APPLICANT: Weiser, Jeffrey N.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
 TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PANITCH SCHWARZ JACOBS & NADEL, P.C.
 STREET: 1601 Market Street, 36th Floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103-2398

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/790,912
 FILING DATE: 29-JAN-1997
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/026,939
 FILING DATE: 23-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Leary, Kathryn
 REGISTRATION NUMBER: 36,317
 REFERENCE/DOCKET NUMBER: 7600-401
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 567-2020
 TELEFAX: (215) 567-2991
 TELEX: 831-494
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1964 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-790-912-3

Query Match 4.1%; Score 223; DB 2; Length 1964;
 Best Local Similarity 22.6%; Pred. No. 6.1e-07;
 Matches 191; Conservative 98; Mismatches 264; Indels 292; Gaps 48;

206 GGHYHYIPKSDLSASELAALAKAGKRNQPSOLYSTASDNNQSVAKGSTSKPA 262
 47 GGHYHYIPKSDLSASELAALAKAGKRNQPSOLYSTASDNNQSVAKGSTSKPA 84
 263 NKSENLOSLKELYDPSAORYSESDGLVDPAKIISRTPNGVAIPRGDHYHPIYPSKLS 322
 85 -YKLSNQNLAEL---PNTGSKNERQALVAG---ASIALGILI-----FAVSKRK 128

```

OY 323 ALEKIRAMPISGTS--TVSTNAKNEVYSSLSNPSSTLTSEKLSADGYIFN 380
DB 129 VKNTVYALVAVMGNGVAVSHALENHL-----LNTN-----IDYELTSGK--LPL 176
OY 381 PKDVEETAAYIYRHDHFIYIPKSNQIGOPPLPNNSLATPSPSLPINFOTSHKEED 440
DB 177 PKESGTYIGYI--KEKRTSDSEVSNO-----EKSAATPT-----KQKV 216
OY 441 GYGFDAIRIIAEDSGFVMSGHGDNHYFFKKDLTEQIKAAQKLEVKTSNGLDSSL 500
DB 217 DYNVTPNFV---DHPSTVQAIOEQTPVSSSTKP--TEVOVVEKPFSTELINPRKEKOSDS 272
OY 501 HEDDYPNAKEMDLDKIEKTAGIMKOYGVRESYVNRKKNALIIYPRGDHHDAPID 560
DB 273 QEQ-----LAHNKLETKKEKES-----PREKGV-----NTLNP0D 305
OY 561 E-----HKFVIGSHSNTELEFPEGVAKKGNKYTGEBELNVNVLKNSFNQNF 614
DB 306 EYLSGQANKP-----ELLYRETIIE--TKIDPOEI-----QENP 338
OY 615 TLANGQKRVSESPPELEKLGIM-LVKLIT-----PDGKYLEVSG 656
DB 339 DLAEQTVRV---KQEGIKGKVEIYRIFSVKKEVSREIYSTSTTASPRIVEGTR 392
OY 657 K--VFGE-----GV-----GNINAFELDQPYLPQGTFTIASKDYEVSYDGTFTVPT 703
OY 393 KTOYIKQOPETGVHKNQVSGALYEPAL-QPELP-----EAVSDKGPPEVOP--TLPE 443
DB 444 AVV-----TDKGET--EVOPESPPTVVSCKEPEQVAPLPRYKGN-----I 482
OY 760 GEIR--LPIRL--NQGTRRAGNKIPTFMANAYLDNOSTIYEVPLLEKENDKRSIL 816
DB 483 EOVPEPPEVETKROGPEKT--EEVPV-----KPEETPVNPEGTEGTSI- 527
OY 817 POFERN---KAOENSKLDEYEEBETSEKY-----EKEKLSSTGSTS 856
DB 528 -QENENVOVPAEESTTSEKVSPTSSSENGEVSNSDSSTSVGESNKPENHNSKENS 586
OY 857 NSTLEVPYDVPQEVAKFAESYGMKLENVLFMDGTIELYLPSEGYIKKNMADFTGEA 916
DB 587 EKYTEEP--VNP-----NEGIVE-----GTSN0TEK 612
OY 917 POGNGENKPSNGVY---STGYENOPTENK--ADSLPEAPNKPYPKPS--TDNG 967
DB 613 PVOAPEETOTSGRIANENGEVSNKPSDSKPVESNOPEKNGTATKPNESGNTSNG 672
OY 968 MLNPE 972
DB 673 QTEPE 677

```

RESULT 5

Sequence 2, Application US/08790912

Patent No. 5976542

GENERAL INFORMATION:

APPLICANT: Welsler, Jeffrey N.

APPLICANT: Plaut, Andrew G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P. C.

STREET: 1601 Market Street, 36th Floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19103-2398

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/790,912
FILING DATE: 23-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,939
FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-401
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELE: 831-494
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2053 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-2

```

Query Match: 4.1%; Score 223; DB 2; Length 2053;

Best Local Similarity: 22.6%; Pred. No. 6,5e-07; Matches 191; Conservative 98; Mismatches 264; Indels 292; Gaps 48;

```

OY 206 GGHYHYIPKSDLSNSELAAKHLAKGNMOPSOISY--STASDNTQSVANGSTSKPA 262
DB 120 GVHYKYVADSELSEE-----KKQVYDIPYVDEDETYILY----- 157
OY 263 NKSENTASLAKELYDSASAPRSESQGVDPKAIIRTEINGVAIPRGDHYHFIYPSKLS 322
DB 158 -YKINSQNOAL--PNTSKNERQALVAG--ASIALAGILI-----FAVSKK 201
OY 323 ALEKIRAMPISGTS--TVSTNAKNEVYSSLSNPSSTLTSEKLSADGYIFN 380
DB 202 VKNTVYALVAVMGNGVAVSHALENHL-----LNTN-----IDYELTSGK--LPL 249
OY 381 PKDVEETAAYIYRHDHFIYIPKSNQIGOPPLPNNSLATPSPSLPINFOTSHKEED 440
DB 250 PKESGTYIGYI--KEKRTSDSEVSNO-----EKSAATPT-----KQKV 289
OY 441 GYGFDAIRIIAEDSGFVMSGHGDNHYFFKKDLTEQIKAAQKLEVKTSNGLDSSL 500
DB 290 DYNVTPNFV---DHPSTVQAIOEQTPVSSSTKP--TEVOVVEKPFSTELINPRKEKOSDS 345
OY 501 HEDDYPNAKEMDLDKIEKTAGIMKOYGVRESYVNRKKNALIIYPRGDHHDAPID 560
DB 346 QEQ-----LAHNKLETKKEKES-----PREKGV-----NTLNP0D 378
OY 561 E-----HKFVIGSHSNTELEFPEGVAKKGNKYTGEBELNVNVLKNSFNQNF 614
DB 379 EYLSGQANKP-----ELLYRETIIE--TKIDPOEI-----QENP 411
OY 615 TLANGQKRVSESPPELEKLGIM-LVKLIT-----PDGKYLEVSG 656
DB 412 DLAEQTVRV---KQEGIKGKVEIYRIFSVKKEVSREIYSTSTTASPRIVEGTR 465
OY 657 K--VFGE-----GV-----GNINAFELDQPYLPQGTFTIASKDYEVSYDGTFTVPT 703
DB 466 KTOYIKQOPETGVHKNQVSGALYEPAL-QPELP-----EAVSDKGPPEVOP--TLPE 516
OY 704 SLAYKMASQTFYFPFHAGDYLRVNPQ---FAVPKGTDLVRYFDEFGHNAVLENNYKV 759
DB 517 AVV-----TDKGET--EVOPESPPTVVSCKEPEQVAPLPRYKGN-----I 555
OY 760 GEIR--LPIRL--NQGTRRAGNKIPTFMANAYLDNOSTIYEVPLLEKENDKRSIL 816
DB 556 EOVPEPPEVETKROGPEKT--EEVPV-----KPEETPVNPEGTEGTSI- 600

```

QY 817 PQRN---KAQNSKIDVEEPTSEK---EKKLSTGTS 856
 DB 601 -QENENVPQAEESTSEKSPDTSSENGEVSNSDSTSYGSEKKEHNDKNS 659
 QY 857 NSTLEPPTVDQEKAKFAESTGKMLENVLFMDOGTIELYLSGVKIKNNADFTGEA 916
 DB 660 EKVTEEP-VNP-----NEGIVE-----GTSNQETEK 685
 QY 917 POGNGENKPSENGKRV---STGTVENOPTENKP--ADSLPEAPNEKPKPKENS---TDNG 967
 DB 686 EYQAEETQNSGKIANTENGEVSNKPSDKPVEESNOPEKNTATKPEKNGTSENG 745
 968 MLEPE 972
 746 QTEPE 750

RESULT 6
 US-08-447-031A-2
 Sequence 2, Application US/08447031A
 Patent No. 5851794
 GENERAL INFORMATION:
 APPLICANT: GUSS, Bengt
 APPLICANT: HOOK, Magnus
 APPLICANT: JONSSON, Hans
 APPLICANT: LINDBERG, Martin
 APPLICANT: PATTI, Joseph
 APPLICANT: SIGMAN, Christer
 APPLICANT: SWITALSKI, Lech
 TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS
 NUMBER OF INVENTIONS: ITS PREPARATION
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Burns, Doane, Swecker & Mathis
 STREET: P.O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/447,031A
 FILING DATE: 22-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/861,804
 FILING DATE: 21-AUG-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/SE91/00707
 FILING DATE: 22-OCT-1991
 APPLICATION NUMBER: SE 9003374-7
 FILING DATE: 22-OCT-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: McGowan, Malcolm K.
 REGISTRATION NUMBER: 39,300
 REFERENCE/DOCKET NUMBER: 012889-006
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1183 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-447-031A-2

Query Match 3.88; Score 205.5; DB 2; Length 1183;
 Best Local Similarity 19.78; Pred. No. 4.8e-06;
 Matches 214; Conservative 139; Mismatches 411; Indels 323; Gaps 50;
 QY 36 NNRYSYDQSSQSSKSNLR--PDQVSOKEGIOAEQIYKLTIDQGYVSHGDHYHYNGK- 93
 DB 195 NNEKSYV-----SDITIKQIOGGQQLDLSTANINT-----GRHSNYISGGS 238
 QY 94 --VPYDALFSEELMKDPNOLKADADI-----VNEVKGITIVDQ--KTYVYLLKDA 142
 DB 239 AITDFEKAFFGSKITVDTNKTIDVTIPQGYSYNSEINKTKITNEQOELEFNNSQAW 298
 QY 143 HADNVRKDEINROKQSHVNDNEKNSNV-----AVARSGRTTN 183
 DB 299 YQEH--GKEEVNGKSFNHYVHNINANGIBGTAKGELVTKLQKODTAPLANKFKLSK 356
 QY 184 DGYVF---NPADIIEDTGNAYIYPHGHYHYIPKSLASSELAAAHLAGKNQPSQL 239
 DB 357 DGSVVKNOKEIEIITLANGI-----ANKALPSGDIILKEIAPRPYFDKKE----- 406
 QY 240 SYSTASDNNQTSYAKGSTSKPANKSENLOSLELYDSFSAQRTSESGLVFPKATIS 299
 DB 407 -YPTFMKDTDNOGYF--TTELEMAKIEKTDV-----SAOK-----YMGSTQVK 448
 QY 300 RTPGVAILPHGDHFIPIYK--LSALEKARAVPISGGSYVSTAKPREVYVSLGSL 357
 DB 449 PTIYFKLYKODDNONTPVDAETKLED-----GTTKVTWMLPEKD----- 491
 QY 358 SSNPSLTTSKELSSADGYIFNPKDIVEETATAYIYRHGHPHYIKS--NQIQPILP- 415
 DB 492 -----KNGKAIYLVKENVAGGEDTPT 514
 QY 416 ----NNSLATPSPSLDI--NPGTSHKHEEDGIGFDANRIIAEDSGFTVSHGDHNYFF 469
 DB 515 GYTKENGVLVYTNTERKIEFTTSISGEKWDKQKODKR--PERVSNLLANG----- 565
 QY 470 KDLTEQIQAQKHLEEVTSNGSLDSHEDDYGNKEMDLDKIE--EKIAGIMK 528
 DB 566 -----EKVAT--LDVTSNMTK-----EFDLPRYDGRKIEYTVT 600
 QY 529 QYGVK-----RESIVVKN-----EKNAIYYPHGDHHAIDPEKPYGISHSNYELFK 578
 DB 601 EDHAKDTTIDNGTITINKTIPGETSATYVKKNMDNNQ--DGRAPREI-----KVLEY- 652
 QY 579 PEEGVA-----KKEGNFY--YGEELINVVNLLKSNTPNNQFT 615
 DB 653 -QDKATGKTAALINESNNMHTHTGLDEKAKGQOVKTYVEELTV-----KGYTTHVDNND 707
 QY 616 LANGQKRVSPFPELEKIKGINLVLLIPDGVLEKSGKVGEG-----VGNIAN 668
 DB 708 M--GNLIVTKKYPETTSISGEKWDKQKODGRPRKVSVNLLADEKVTLDVTSNTN 765
 QY 669 FELDQPILP-----GQTEFYTIAS--RDYPEVSVDGFTVPTSLAYKASQTIYPFHAG 721
 DB 766 WKYEFKLPKIDBEKRIEYIVTEHDVADY--TIDNGI-----TINKTKTPG 810
 QY 722 DTYLRV-----NPOFAPVPGTDALVAFDE-----FHGNAYLENNYKV 759
 DB 811 ETSATVTKNMDNNOGKRPTEIKVELYODGKATGTAALINESNNMHTHTGLDEKAKG 870
 QY 760 GEIKLPKPLNQ-----GTRTAGNKIPTYPMANAYLDNOSYIYVPLILEENQDIX----- 812
 DB 871 QQVYVTEELTKVKGYYTHVDNMGMLIYTNKTPETTSISGEKWDKQKODGRKPER 930
 QY 813 --PSIILPQFRKNA-----QENSKLDE--KVEEPTSEKVEKELSETQNST 855
 DB 931 VSVNLLANGKERTLDVTSNTKYEKDKLPKIDBEKRIEYIVTEHDVADYVTDYDNGTIT 990
 QY 856 SNSTL--EEVPTVD-----PVQEKAKFAESTGKMLENVLFMDOGTIELYLP 900
 DB 991 TNKTYPETSATYVKKNMDNNOGKRPTEIKVELYODGKATGTAALINESNNMHTHTWG 1050

QY 901 GSEVTKMMADFT-GEAPOGNGENKPSBNGKSTGTVEONPTENKPADSL-PEAPNEK-- 956
 DB 1051 LDEAKGGOVATYVDELTKVNGITTHVDNDGMLTYTNKTPPKPKPIPEPKDTP 1110
 QY 957 PVKPEMS 963
 DB 1111 PTKPDHS 1117

RESULT 7
 US-08-296-791-6
 Sequence 6, Application US/08296791
 Patent No. 6245337

GENERAL INFORMATION:

APPLICANT: St. Geme III, Joseph W.
 APPLICANT: Falkow, Stanley
 TITLE OF INVENTION: Haemophilus Adherence and Penetration
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/296,791
 FILING DATE: 25-AUG-1994
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Trecartin, Richard F.
 REGISTRATION NUMBER: 31,801
 TELEPHONE: (415) 781-1989
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249

INFORMATION FOR SEQ. ID NO.: 6:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1848 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 US-08-296-791-6

Query Match 3.88; Score 203; DB 4; Length 1848;
 Best Local Similarity 19.08; Pred. No. 1.4e-05;
 Matches 259; Conservative 152; Mismatches 447; Indels 508; Gaps 64;

2 KFSKRYIAAGAVYVSLCAVALNHR--GKNKNNKRSYDGS 45
 DB 5 KFKNFAL--TAAVLTPTTEALVDDVDYOLFDFNAENKGFSGATNVE 55
 QY 46 QSSOKSENT--TPDOYSOK--GLOAQOYIKITDQYV--TSHGD--HYHYNG 92
 DB 56 VRDKKNSLGSALFNGIPMIDFSVVDKRAIATLVNPOYVGVHNGVSELHFGWNG 115
 QY 93 KVPYDALPSELKMDPNYOLKADIVNEVKGITIIIVDGKTYVYLDAADANVR--TK 150
 DB 116 N--NNNGNAK--SHRDVASE--ENRYTYVEKNNEPTENVTSTFK 153
 QY 151 DEINROK--HYKDEKVNVA--VARSGRYTNDQYVNP-- 191
 DB 154 EEDQAKRRREDYPRDKTEVETVPIEASTANNKKEIINSDY--PAVRLGSGTGF 210
 QY 192 --DIIEDTGNAIYVPHG-- 207

DB 211 IYKRSRYQLILTEKROGMLLRWDVGDNLLEVNAAYYGIAGTYKVNHNGLIGF 270
 QY 208 --HYHYRSDLSASELA--AKAHLAGNMPSPQ 238
 DB 271 GNSKEEHSDDPGILSODPLTNVAVLGDGSPFLYDREKGMFLGSDYDWAGNKSWMQ 330
 QY 239 --LSYS--TASDNTQSV--ANGSTSKPANKSENLKELDYDPSA 281
 DB 331 EWNITKEPAKTIYQOVSAGSLISNTQYWOAGTSTITGGEPPLSDLTQCKDPNH 390
 QY 282 QRY--SESDGLVDPDAITISRTPN--GVAIPRGD--HYH 314
 DB 391 GKSITLKGSGILNHNIDGAGGLFPEGDYEVGTSDSTWKGAGVADAGKYTVKWH 450
 QY 315 FIPYSKLS--ALEEK--IARNAVISGTS 339
 DB 451 NPKYDRLAKIGKGLVYEGKKNGLKVGDPGYTLKOKDANKVOAFSGVGVSRST 510
 QY 340 TVSTNAK--PNEVYSSL--GSLSSNPSL--TSKELSSASD 375
 DB 511 LVLNDKQVDPNSIYFSGRGRDLNNGNSLFFPHIRNIDGARYVNNMTVTSITITGE 570
 QY 376 GYIFNPDIYEETATAYIVRGHFR--YIPKSNQIQOPLPNNSLATPS 423
 DB 571 SLITNP--NTTSYNIEAQDDHPLRISRYRQLYFQDNR--SYTLKKG--ASTR 622
 QY 424 PSLPINFQTSHEKEEDGYGDA--KRILEDSSGYMSGDNHFFKDLDEQ 477
 DB 623 SELPONGESENEMLYMGRTSDAKRVNMMHNNRNGF--NGVF--GEE 670
 QY 478 IKAOKHLEEVKTSNOLDISLSEHODY--PGNAKEKDLKKEETIAGIMKOYKRE 535
 DB 671 TKARON--GLINTYFNG--KSDQNFLLGNGNLNDL-- 704
 QY 536 SIYVKKERNAIYPHGDHHDPIDEHKPYGIG--HSHSNTELFKEEGYAKT-- 586
 DB 705 --NVEKGLFLSGRPTPHARDI--AGISSTKDPHTENNEVEVDEDMVNRFXA 755
 QY 587 --BGN-KYTYGEELTNVNNLKNSTFNNONTTANGOKRVFSFPELEKLGIML 640
 DB 756 TTMNVNTNASLISGR--NNAITISNTASN--NAOVHIGY--RTGJTV 798
 QY 641 VR--LTPDGKVLKESGKYGEGV--GNI--ANPELDQPLPQOTFEYITASK 688
 DB 799 VRSDYTGTYCHNSNLEKALNSFPNLTGNVLTENASTYLGKALIFG--TIQSI 853
 QY 689 DYPEVSDGTFTVPTSLAYKMASQITIFPPHAGDTYLRVNPQAFAPRGDALVRVDEFH 748
 DB 854 GTSQVN--LKENSHWHLTGSNVNQLNLTNGHILNANQDANKVTTYNTLTVSLS 907
 QY 749 GNA--YLENNKYVEIKLPT--FLNNGT--TRTGNKILPV 783
 DB 908 GNGSFYTVVDFTNKSKRVVYVNSATGNFTLOVADKTGEPHMLTLFLDASNAVRNMLEV 967
 QY 784 TFMANAYLDN--OSTYIVEPILKEKNOC--DKRSILPQF 819
 DB 968 T-LANGSVDRGAMKYKLRANNGRIDLYNPEYERKNQYVDTNTTTPMDIQADAPS-- 1021
 QY 820 KRNKAQENSRLDEKVEEP--KTSEKVEKER 847
 DB 1022 --AQSNEELIARVETPVPPAPATSAIASQEPETRAPETAQAPAMEETNTASTTAP 1077
 QY 848 LSETGNSTSLTEPVTVPVOEKVAKFASYSKMLENLVFNNDGILIELXPGEYIK 907
 DB 1078 KSDPATOTENPNSESVS--ETTKVAVENPQENETVAKNQEATEPTPOGGEYAKE 1132
 QY 908 NMADF--TGEAPOGNGENKPSBNGKSTGTVEONPT--ENKPADSLPEANER 956
 DB 1133 DQPTVEANTQNTAKTQSEGTETETQ--TAKKSEPTSVYSENOPEKITYSQSTEDK 1187
 QY 957 PV--KPNSTDNGLNPEGNVSDP--MLDPALAPAVDPVOE 996
 DB 1188 VVVEKEKAKVETEETQKAPQVTSKEPPKQAPAEVPPDTJNAEE 1233

RESULT 8
Sequence 6, Application PC/TUS9510661A

GENERAL INFORMATION:

APPLICANT: Washington University, et al.
TITLE OF INVENTION: Hemophilus Adherence and Penetration Protein

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PC/TUS95/10661A

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791

FILING DATE: 25-AUG-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: PP-59941/REF
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

INFORMATION FOR SEQ. ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US95-10661A-6

Query Match 3.88; Score 203; DB 5; Length 1848;

Best Local Similarity 19.08; Pred. No. 1.4e-05;
Matches 259; Conservative 152; Mismatches 447; Indels 508; Gaps 64;

QY 2 KFSKKYIAGSAVYSLCAVYALNQRS-----QEKKNKRYSYDGS 45
DB 5 KFKLNFTAL-----TVAVLTPTTEALVBDVDYQIFRDFAEKRGFSVGATNVE 55
QY 46 QSSORSEML--TPDQVSOKE--GIOAEQVIKITDQGV-----TSHGD--HYHYNG 92
DB 56 VRDKKNSIGSALNGIPMIDFSYVDKRIATILVNPQYVGVKHSVGSSELFNGING 115
QY 93 KVPYDALFSEELMKDPYQIKADIVNEVKGITIIIVGKTYIYLDAAHADNVR--TK 150
DB 116 N-----ANNNGAK--SHRDVSS-----ENRYTYVEKKNPFTEVNTSPTK 153
QY 151 DEINRQOE--HYKDEKVSNA-----VARSGRYTNDQYVFNPA-----191
DB 154 EEOAQKREDYMPRLDKFTTEVAFLASTANNKGEYNSDY--PAFVRLGSGTQF 210
QY 192 -----DIIEDTGNAYIYPHGG-----207
DB 211 IYKSGSRQQLITKEDQGNLRMDVYGGDNLELVGNAYITGIGAGTPTKYVHENGILGF 270
QY 208 ---HYHTIPKSDLSASELA-----AAKAILAGKRMQPSQ 238
DB 271 GNSKEHSDPGLISDPLNTYAVLGDGSGPLFYDREKQKWLFLGSDVFAAGTNKKSMD 330

QY 239 -----LSTSS--TASDNTQSV--AKGSTRPANKSENLOSILKELYDSPA 281
DB 331 EMNITKEHFAKTIQOYANGSLIGSNQYTMQANGSTETIGGSEPLVDLGDKNH 390
QY 282 QRY-----SESDLVETDPAKIIISRTPN-----GVAIPRGD---HY 314
DB 391 GKSITLKSGTLTNNHIDQAGGLFEFGDEYVKOTSTWKGVGAVADGKYVTMKNH 450
QY 315 FIPYSKLS-----ALEEK-----IARVPISSGGS 339
DB 451 NPKYDRILAKIGKGLVYEGKGNKNGELKVGDGTVILKQKADANNKVAQPSGVGSGST 510
QY 340 TVSTNAK--PNEVYSSL--GSLSNPSSL-----TSKELSSASD 375
DB 511 LVLWDDKQVDPNSIYFGRGRIDLNGSLFPHDRINIDGQARVYNNHMTSNTITGE 570
QY 376 GYIFNPDIYETATAYIVRGDHR-----YIKSNQIGQPTLPNNSLAPPS 423
DB 571 SLITNP-----NTITSYNEAQDDHPLRISIPRQLYFNQDNK--SYTYLKG--ASTR 622
QY 424 PSLPINTGTSHEKHEGQYGFDA-----NRIADESGFYMSGDHNYFFKDLTREQ 477
DB 623 SELPONGSESNEMLYMGRTSDAAKRYNNHINNERANGF-----NGYF-----GEE 670
QY 478 IKAOKHLEKVTSHNGDLSLSHEDY--PGNAKEMDLDKIEEKIAGIKQYGVARE 535
DB 671 TKATON--GRANTFNG--KSDONRELLTGCTYNDL-----704
QY 536 SIYVNERKNAIYPHGDHHDAPIDEKRPVIG-----HSHSNTELPREGVAK--586
DB 705 ---NVEKGLFLSGRPFPARDI-----AGISFKKDPHTENNEVVEDWIRNPKA 755
QY 587 -----EGN-KYITGEELTNVNLKNSFNNOHTIANGQKRVSSFPPELEKIGIML 640
DB 756 TTMNVTGNASLYSGR--NVANITSNTASN-----NAQVHIG--KTGDYVC 798
QY 641 VK-----LTPDGKYLEKYSKGFEGEV--GNI-----ANFELDQYLPQGTFFKTIASK 688
DB 799 VRSDYTGVTCHNSLNSKALNSFNPTLRGNVILTEASLTIGKANLFG-----TIOSI 853
QY 689 DYPEYSYDGTETVPSLAYKMAQSOTIEYFPFAGDTYLRVNFQFVAPKGTALVRYFDEFH 748
DB 854 GTSQVN-----LKNSHMHLTGNSNVNQNLNLNGHILNQNQDANKTYTNTLVNLS 907
QY 749 GNA-----YLENNYKVGEEKLP-----PLNGT-----TRTAGKLPY 783
DB 908 GNGSFYTWVPTNKNKSNRVVYVKSATGNTLQVADKGTGEPHNEILTFEDASNATRNMLEV 967
QY 784 TFMANATLDN-----OSTYIYEVPILEKENOT-----DKPSILPOF 819
DB 968 T-LANGSYDRCAMKTKLRANNGRIDLINPEVEKRNQYVDTNITTPNDIQDAP8-----1021
QY 820 KRNTAOKNSKIDKEVEEP-----FTSEKEREK 847
DB 1022 ---AOSNNEIARVEFPPAPATASALASEQEPETPAPTAPAMEETWTANSTETAP 1077
QY 848 LSETGNSTNSLTLEVPYDVQERVAKFAESYGMKLENTLPNDGTIELTLPGEVYIK 907
DB 1078 KSDPATOTENNSSEVPS-----ETTERVAENPQOENETAKNBOENATEPTPOGEVAK 1132
QY 908 NMADE-----TGEAPOGNGENKPSENGKVSCTGVENOT-----ENKPADSLPEAPK 956
DB 1133 DQPTVEANTQNEATQSGKTEETO-----TAEKSEPTSEVYSNENQPEKTVQSGSTEDK 1187
QY 957 PV-----KPNSTNGMLNPEGNGSP--MDPLLEAPAVDPQOE 996
DB 1188 VYVEKEEKAKYETETQKAPQVTSKEPKQAEPAPEEVPDTNABE 1233

RESULT 9
US-08-961-083-4
Sequence 4, Application US/08961083
Patent No. 6159469

Db 407 L-----AKIGKGLTINGVNOGOLKVDGTVILNQKADADKVOAFQVGI 454
 QY 351 VSSLSGSSNPSLTSTKSLSSASDG--YIFNPKDIYEE-----TATATVNRG-DHFHY 402
 Db 455 VSGRGTLVLLNSNINPDNIYEFGRGRLDANNDLTFEIRNVDCGARIVNTHAST 514
 QY 403 IPRSNIOGPTLPNNLAPPSLPINPGTSHEKHEEDYG--FDANRIIAEDSEGFVMS 460
 Db 515 I-----TLGKSLITPNLSLVH--SIQNDYDEDDIYTYRPRRPPOGKDIYKN 563
 QY 461 H-----GDHNYFFPKDLTEEOIKAAOKHL-----BE 487
 Db 564 YRYALAKSGRLNAPMPENGVANNDMIMWGYGEARNAHMKNNRIRIGDCGFDEE 623
 QY 488 VRTSHNGDLSL-----SSHEODYPGNAKE 511
 Db 624 NGRGHGALTLNNGSAONRELLTGTGANLCKISTOGVLLSGRPTPHADPVKSSA 663
 QY 512 MKD--LDKIEEKINAGIMQYKRESIVYNNKKNALITPHGDHHDPI-----DEHK 563
 Db 684 KDAHFSKNEVVEEDDMINRTFAAEIIVNOSAS-----FSSGRVDDITANITATDN 738
 QY 564 PVGIGSHSNYELFKPEEGVAKKEGNKV-----YTGEELTVNVLKSTFNQNTFLAN 618
 Db 739 KVALGY-----KNGDEVCRSDYTGVTGNTGN-LSKALN-----773
 QY 619 GQRVSPSPPELEKRLGIMLVKLITPDGKLVKYSKAVGEGCNANFELDQYLP 678
 Db 774 ----SFDXATKINGNVNINQNALVL--GKA--ALMGKIOGGNSRVSINQSKHLNG 824
 QY 679 QTKYTIASD-----YREVSYGTFVPSIAYKASQITFP 717
 Db 825 DSOVHNLSLADSHILNNSADQASANKYTKIKINLSGNGHFYTLDLANLADKVLVE 864
 QY 718 FHAGDYLVRYNPQFAVP--KGTDL--VRFDEFHGNATYENNY-FVGEIKLPIPLNQG 772
 Db 885 ASAGHQLAHQNTGEPNOEGDLFDASSVQDSRLFVSLANTYDGLALRTI-KTENQ 943
 QY 773 TTRFT---AGNKIPVTFMANAYIDNOSTYIVEVPILEKENQD---KPSII---PQF 819
 Db 944 ITRLYNPYAGNGRPVAPSPANTASQ-----AQKATQTDGAQIAKPNQNTVAPPSP 996
 QY 820 KKKKAQENSLDKVEPTSEKVEKEKISEGNSNSTLEVPYVDVQEVAFPA--877
 Db 997 QANQAE--KQAKAQOQVROQAEEKVARQDEAKRKAALIAQOEBAARAALAAK 1055
 QY 878 -----ESYGMKLENYLFMDGTIE---LYLQSEVYIKKNMAFP---TGEAPQ 918
 Db 1056 QKAERKARELARQKAEBSHOAKPKRRRRRALLPRPAPVSLDIDYADANSESSI 1115
 QY 919 GNGENPSENGKYSTGTVENOPTENKPADSLPE-----APNEKVPKPNSTDNQW 968
 Db 1116 GNLARVPYIRGRILINDYEBIPLLELEDEAEERROATGPHSKSRMRRAISSPSSDEDA 1175
 QY 969 LNFEGNVGSPMLDPALEAPAVDVOEKLEKFTASTYGLDSDVITN-----MDGTI 1020
 Db 1176 SESVSTSDKRPQDNTLHEVETAGLQPRAPAOPTQAAAQADAVSTNTNSALSDAASQ 1235
 QY 1021 ELRLPSGEVYIKKNLS 1035
 Db 1236 SIILDTGAVYITRHA 1250

RESULT 11
 US-08-904-263A-4
 : Sequence 4, Application US/08904263A
 : Patent No. 6015869
 : GENERAL INFORMATION:
 : APPLICANT: LINDAHL, GUNNAR
 : APPLICANT: STAHLHAMMAR-CARLEMALM,
 : APPLICANT: MARGARETHA
 : APPLICANT: STENBERG, LARS
 : TITLE OF INVENTION: PROTEIN RIB, A CELL SURFACE PROTEIN THAT

TITLE OF INVENTION: COMBENS IMMUNITY TO MANY STRAINS OF THE GROUP B
 TITLE OF INVENTION: STEPTOCOCUS: PROCESS FOR PURIFICATION OF THE PROTEIN,
 TITLE OF INVENTION: REAGENT KIT AND PHARMACEUTICAL COMPOSITION
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
 STREET: PO BOX 747
 CITY: FALLS CHURCH
 STATE: VA
 COUNTRY: USA
 ZIP: 22040-0747
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/904,263A
 FILING DATE:
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: SVENSSON, LEONARD R
 REGISTRATION NUMBER: 30,330
 REFERENCE/DOCKET NUMBER: 552-119P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 205-8000
 TELEFAX: (703) 205-8050
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1231 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-904-263A-4
 Query Match 3.4%; Score 186; DB 3; Length 1231;
 Best Local Similarity 20.3%; Pred. No. 0.00012;
 Matches 240; Conservative 145; Mismatches 416; Indels 384; Gaps 66;
 QY 39 VSTVDSQSSQSENTPPOVSQKREGIOAEQIYKITDQGYVTSNG---DHYHYNGRV 94
 Db 35 ISFLGG--FTQGFNISTGTVFPAEVISSAVTLN-TNNTKNQNGRAVITDLYDVANGKI 91
 QY 95 -PYDA-----LFSELLMKDPNT--QLKADIVNEVGGY-TIVVDGTY-----135
 Db 92 DPLQITLNSPDKAQAYIRQNGNYFTQPSLTVGAASINTVLTQDSPHTRPGQYD 151
 QY 136 -----VYLKQAAH--ADNVRK-----DEINRQK-----QEHVKNKENVSNYA 172
 Db 152 IINVSITITNSSALRDKIDVKKKKAEDPKWDESSRKLVLISLDIDITDIDNNPKTQSDIA 211
 QY 173 VASQGRYTTNDGYVNPADILIEDTG-NAYIVPHGCHYHYPKSDLSASELAANKAHLAG 231
 Db 212 -----NKITEVTLKELIVR-----IPDAD-----KNDPAG 238
 QY 232 KNNQPSQLSYSTASDNNTQSAKGSSTKPAKSENULQSLK---ELTSPSAQKRSSESD 288
 Db 239 KD-----QGVNMGTPRAEDSIGMLPDLKQTVAFETP-VDTATPBD 280
 QY 289 GLVFDPAKIIISRTPNG-----VAIPHGDHYHIFIPYSKLSALEKIAKAVPISGIG 338
 Db 281 ---KPAKVAVVTPDQSKDTVDYTVAVVDPRTDADKNDP-----AGKD 319
 QY 339 STVSTNAKPNVSVSLGSSNPSLTSTKE--LSASDG-----YIFNPKDIYEE 387
 Db 320 QGVNMGTEP-KAEDSIGMLPDLKQTVAFETPVDATGDKRAKVAVVTPDQSKDTVDY 378
 QY 388 TATATYIVR-HGDHFHTIPYSNQT-----GQPTLPNNSLATSPSLPIPIPGS 433
 Db 379 TVYVVDPRTDADKNDPAGKDOQVNVGETPKAEDSIGMLDLPFGT--TAAFEPTVDATP 436
 QY 434 HEKHEDDYGFDANRIIAEDSEGFVMSHGDHNNYFFPKDLTEEOIKAAOKHLEEVTSN 493

```

Db 437 GDK-----PAKVVYTPDGS-----KDVVDVYKVVDPRTDADKNDP 474
QY 494 GJLSL-----SHEDYPRGAKEMDLKRIE-----EKIAGIMQY--GK 533
Db 475 GKQOVNVGEPRAEDSIGN--LPDLPKGTVAETPVDTAPGKPAVVVYTPDGS 531
QY 534 RESIVVAKENALITYPHGDHHAIDPEHPRVIGSHSHSYELFKPEGVAKKGNKYV 593
Db 532 DTVDTYK-----VDPRTDADKNDPAGKQOVNVG-----ETPKAEDSIG----- 572
QY 594 GEEELVNVNLKST--FNNONFTLANGOF--RVSEFPELEKLGIMLVLI----- 644
Db 573 -----MLPDLPKGTVAETPVDTAPGDPKPAVVVYTPD--GSKDTVDVYKVVDPRTD 625
QY 645 -----TPDGKLEVSGKV--FGEGVGNIANFELDQYTLGQTKYTIASADPEVSDGT 688
Db 626 ADKNDPAGKQOVNVGEPRAEDSIGNL-----PDLPRGT--VVA----- 663
QY 699 FTVPTSLAY--FMASOTIYEPHAGDTYLRVNPQFAPVPGTDALVRVD-----EFGH 750
Db 664 FETPVDTAPGDPKPAVVVYTPDGSKDT-----VDVYKVVDPRTDADKNDP 710
QY 751 AYLENNYKVEIKLP-----IPKLNQGT-----FTAGNKIPVTEMANAYLDNOS 795
Db 711 AGKDOOVNVGEPKADSDIGNLPDLPKGTVAETPVDTAPGDK--PAKVYVYTPDGSND 769
QY 796 TYIVVERPILEKENQDKPSILPQFKRKAQENSKLDERV--EEPTSEKVEKEKLESG 852
Db 770 TVDVYKVVDPRTDADK-----NDPAGKQOVNVGEPRAEDSI--GNLPDL 815
QY 853 NSTNSTLEVPYVDPVOEKVAFKASY--GMLENVLENMGTIELILPSGVITKMA 910
Db 816 KGTVAETPVDTAP--GDPKPAVVVYTPDGSK-----DTVDVYKVVDPRTDADKNDP 869
QY 911 DFTGAPQGN--GENKSENGKYSTGVENQPTENKPADSLP--EAPNEKPYV-----PE 961
Db 870 ---GKQOVNVGEPKRAED--SIGNLPDLPKGTVAETPVDTAPGDKPAVVVYTPD 923
QY 962 NS-----TDNGMLNPEG-----NVGSDPMLDPALEAPV-----DP 993
Db 924 GSKDTVDVYKVVDPRTDADKNDPAGKQOVNVGEPKADSDIGNLPDLPKGTVAETP 983
QY 994 VQ-----EKLEKFTASYGJGDSVIFNMGTIELRLPSGEVAKKN 1033
Db 984 VDPATGDPKPAVVVYTPDGSKDTV--DVYKVVDPRTDADKND 1025

```

RESULT 12

US-08-596-291-3
Sequence 3, Application US/08596291
Patent No. 5821075

GENERAL INFORMATION:

APPLICANT: GONZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESON-WELSH, LENA
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,291
FILING DATE: 09-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: L0461/7000
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
TELEX: 92-1742 EZEKIEL
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2465 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-596-291-3

```

```

Query Match 3.3%; Score 178.5; DB 2; Length 2465;
Beet Local Similarity 18.7%; Pred No. 0.0011;
Matches 186; Conservative 152; Mismatches 396; Indels 261; Gaps 43;

QY 45 SOSQKSENILTP-----DVSQKREGIOABQIVIKITDQGYVSHGDHYHYNGKVPYDAL 99
Db 1434 SPTSKEHYVTPQCTLSDDNAGQGEKVKKTQYVMDSEFVE-----ENT 1479
QY 100 ESEELIMDPNTQL--KDAIVNEYKGGYIIRKVDKTYVYLKDAHADNVRKTD--EI 153
Db 1480 FEVKLFKNSSGGLFSRSDNLIPEQINASIYRV--KILFACQPAESKIDVGVILNV 1537
QY 154 NKQKQHVNDKKNVSNVAVARSQGYTNDGYV--NPADIIDETGNAYIVPGGHYH 210
Db 1538 NGASLGLSQQYI-----SALRGTAPEVFLLCRPPEVLEIDITALLTPQLSPAQ 1589
QY 211 YIPKSDLSASELAAKAHLAGRMOPSOVSSTASDNTQVANGSTSPRKNSENTQS 270
Db 1590 VLPNSKSS-----QPSCYEGSTSSDENMSDKSKCKSPRR----- 1629
QY 271 LKELTDSQAQRYSESDGLVDPKAIISRTPNCAIIPAGDHRIPIYSKLSALEKIR 330
Db 1630 -----DSYSDSGSGEDLVYAPANISNYS-----SALHQTLSN 1665
QY 331 NY-----PISGTGSVSTNAKPNEYVSSLGSLSSNPSLITTSKELASDGYIENP 381
Db 1666 MYSQASHHEAPKSOEDTICTMFTYTPQKIPNKPFEEDSNPSLP-----PDMAPOGSTOP 1720
QY 382 KDIVETATAAYIVRAGDHEHYIPKSNQGOPTLPNNSLATPSPSLPINTGSHKHEEDG 441
Db 1721 QS--SSASSSSDKY--HJHH-----ISEPTROEN--WPLKLN-----DLEHNLBD 1760
QY 442 YCFDANRIIAEDS-----GFVMSGDHNNHYFFKDLLEBOJKAOK-----HLEVKTS 491
Db 1761 FELEVELELLTLTKSEKASIGFTVTKGNORIGCYVADYIODPAKSGRKLPGDRLKV-- 1817
QY 492 HNGLSLSHEDDYGNKAKEMDLKRIEKLKAGIMQGYVR-----ESIVVKEK--NA 545
Db 1818 -NDVYTNKTHIDAVNLAALBAASKYVLYIGRPRITONTNVASFPATGKRLTKNKEELGFS 1876
QY 546 IYIPGHDHHAIDPEHPRVIGSHSHSYELFKPEGVAKKGNKYVYGEELTVNVNLK 605
Db 1877 LCGGDSLYQVYISDINP-----RSVAIAIGNV-----QLDVIYHYVN 1915
QY 606 NSTFNNQFTLANGOKRYSFPPLEKELGIMLVKILTPDGKLEVSGVAPBEDGN 665
Db 1916 G--VSTQGTLEEVNRALDMSL--PSLVTKATRNNDP--VVPSSK--RSASVAPKSTKGNGS 1969
QY 666 IANFELDQPYL--PGQTFYTIASKDYPEVSYDGTFTVPTSLAYKMAQOTIYEPHAGDTY 724

```

DB 1970 YSVGSCSPALTPNDSPS-TVAGEINEIS-----YKGCSTY 2007
 QY 725 -LAVNPOFAVPKGT-----DALVRFDFEGHNAVLENNYKGEIKLP 766
 DB 2008 QINGSPMLTLPKESYIOEDDIYDSQAEVYISLDVYDEAONLNNEN-----2057
 QY 767 PKLNOGTTFRAGNKI-PYTFMANAYLDNOSTIYEV---PILKRENOTDPSILPOKRN 822
 DB 2058 -----AAGSCGPGTLKANGKLSERTEDTDCGSLPEFTTATMNGCEECE 2108
 QY 823 KAOENSLDKVEPEKTS-----EKVEKELSETGNSNSTLEEVPTVDPV 869
 DB 2109 KYKSES-LIOKPOEKTDDEITWGNDELPIERTNHEDSKDSFLNDELAVLPV---2164
 QY 870 QERYAFAESTYGMKLENVLENMDGTIELYPSGEVIRKRNADFTGEAPOGNGENKPSENG 929
 DB 2165 --KVLPSGKYTGANKLSVIVLNGLLDQIGPSKEL--ENLQ-----ELKPLDQC 2209
 QY 930 KYSTGTVENOPTENKPADSLPEAPNEKPYKPESTDN 966
 DB 2210 LI--GOTKENRRRRYKNILPYDATRYPLDDEGYIN 2244

RESULT 13

US-09-100-804-3
 Sequence 3, Application US/09100804
 Patent No. 6066472

GENERAL INFORMATION:

APPLICANT: GONEL, LEONEL JORGE
 APPLICANT: SARAS, JAN
 APPLICANT: CLAESSON-WELSH, LENA
 TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
 TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
 TITLE OF INVENTION: TYROSINE PHOSPHATASES
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
 STREET: 600 ATLANTIC AVENUE
 CITY: BOSTON
 STATE: MASSACHUSETTS
 COUNTRY: USA
 ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/100, 804
 FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/596,291
 FILING DATE: 09-AUG-1996
 APPLICATION NUMBER: US 08/115,573
 FILING DATE: 01-SEP-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/09943
 FILING DATE: 01-SEP-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: GATES, EDWARD R.
 REGISTRATION NUMBER: 31,616
 REFERENCE/DOCKET NUMBER: L0461/7003
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-720-3500
 TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2465 amino acids
 TYPE: amino acid

TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-100-804-3

Query Match: 3.3%; Score 178.5; DB 3; Length 2465;
 Best Local Similarity: 18.7%; Pred. No. 0.0011;
 Matches 186; Conservative 152; Mismatches 398; Indels 261; Gaps 43;

QY 45 SOSOSKSENITP-----DVSQKEGLOAEQIVIKIDOGVYTHSGHGHYHYNKRPYDAL 99
 DB 1434 SPTSKREHVPTPOCTLSIDNAOQGEKRYKTQVADYSFVTE-----ENT 1479
 QY 100 FSEELMKDQPNYOL--KQADIVNEKGGYIIKVDGYYVYLDAHADAVRTKD--EI 153
 DB 1480 FEVKLFKNSSGCLFSPSRREDNLPEQIMNSTVY--KLEPAGPAAESGKIDGVILKY 1537
 QY 154 NROKREHYKDNENYNSNVAVARSQGRYTTNDGYF--NPADIIETGNAYIYPHGHHY 210
 DB 1538 NGASLKGISQOEVI-----SALRGTAPEVFLLCRPPGYLPEIDTALPLQSPAQ 1589
 QY 211 YIPKSLASSELAAVAHLAKRMQPSQLSYSTASDNTQYAKGSTKPAKSENLOS 270
 DB 1590 VLPNSSKDS-----OPCYOQSTSDENEMSKCKOCKSPSR--1629
 QY 271 LKELVDPSPAGRYSDGVDFDPKIIISRTPGVAIPHGDHYHFIYSKLSALEKRIAR 330
 DB 1630 -----DSTSDSSGSGEDLVTPANISNTWS-----SALHOTLSN 1665
 QY 331 MV-----PISGTGVSTNAKPNVYVSSLSNPSLSLTSELSASDGYTFNP 381
 DB 1666 MVSQAQSHHAPKSOEDTICTMFTYIPQKIPKREFEDSNSPLP-----PDMAPGQSYOP 1720
 QY 382 KDVEETATAYIVRHGDHFYIKSNQIGOPTLPNNSLATPSPSLPINFCTSHKEHEOG 441
 DB 1721 QS--ESASSSMQY--HIHH-----ISEPTQEN--WPLRN-----DENHLED 1760
 QY 442 YGFDARKIIMEDS-----GFVSHGDHNYFERKDLTEQIKRAOK-----HLEEKTS 491
 DB 1761 FELEVELLITLISEKASLGFYTKGNRGICVYHDVIOPAASDGLKAGDRILKY--1817
 QY 492 HNGLDLSHEDQYPGNAKEMKDLKRIEKGIMKQYVKA--ESIVYKREK--NA 545
 DB 1818 -NDTDTNMTHTAVNLLRAASTKTVLYIGRVIRIQTNTVVASFGHKLCKKEELGFS 1876
 QY 546 IIPPHGDHRAHDPIDEKRPVIGISHSHNYELFPEBGAANKGKRYTGEELNVAVLK 605
 DB 1877 LCGGHSLYQVYISDINP-----RSVAIEGNL-----QLDVIHYVN 1915
 QY 606 NSTFNNOFTLANGOKRVSPFPELEKIGIMLVYLIPDGKYLEKVGKVFGEVGN 665
 DB 1916 G--VSTQMTLEEVNRLDMSL-PSLVKATRDLP--VPSST-RSAVASAPSTKNGS 1969
 QY 666 IANFELDQPL-PGQTFKYIASKDYEVSDTFTVPTSLAYKMSQTLFYPFHAGDY 724
 DB 1970 YSVGSCSPALTPNDSPS-TVAGEINEIS-----YKGCSTY 2007
 QY 725 -LAVNPOFAVPKGT-----DALVRFDFEGHNAVLENNYKGEIKLP 766
 DB 2008 QINGSPMLTLPKESYIOEDDIYDSQAEVYISLDVYDEAONLNNEN-----2057
 QY 767 PKLNOGTTFRAGNKI-PYTFMANAYLDNOSTIYEV---PILKRENOTDPSILPOKRN 822
 DB 2058 -----AAGSCGPGTLKANGKLSERTEDTDCGSLPEFTTATMNGCEECE 2108
 QY 823 KAOENSLDKVEPEKTS-----EKVEKELSETGNSNSTLEEVPTVDPV 869
 DB 2109 KYKSES-LIOKPOEKTDDEITWGNDELPIERTNHEDSKDSFLNDELAVLPV---2164
 QY 870 QERYAFAESTYGMKLENVLENMDGTIELYPSGEVIRKRNADFTGEAPOGNGENKPSENG 929
 DB 2165 --KVLPSGKYTGANKLSVIVLNGLLDQIGPSKEL--ENLQ-----ELKPLDQC 2209

QY 930 KVTGTVENQPTENKPADSLPEARENEKPVKPESTDN 966
 : | : : | | : |
 Db 2210 LI--GQTFENRRKNRYKNILPYDATRVPPLGDEGGYIN 2244

RESULT 14
US-08-296-791-5

sequence 5, Application US/08296/91
; Patent No.: 6245337
; GENERAL INFORMATION:
; ADDITIONAL: 54 CWS TIT 106000

APPLICANT: Falkow, Stanley

TITLE OF INVENTION: Baemop

TITLE OF INVENTION: Protein

CORRESPONDENCE ADDRESS: NUMBER OF SEQUENCES: 9

ADDRESSEE: Flehr, Ho

STREET: 4 Embarcadero

CITY: San Francisco
STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: F10000

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC

SOFTWARE: PatentIn K
CURRENT APPLICATION DATA

APPLICANT AFFIDAVIT DATA
APPLICATION NUMBER:
;

FILING DATE: 25-AUG-

CLASSIFICATION: 435

NAME: Trecartln, Ric

REGISTRATION NUMBER: ;

REFERENCE/DOCKET NUMBER: A-59941/

TELEPHONE: (415) 781-

TELEFAX: (415) 398-3

TELEX: 910 277299

1 - INFORMATION FOR SEQ ID NO

2 SEQUENCE CHARACTERISTICS

LENGTH: 1702 mm

TYPE: amino acid

! JOPULOGY: unknown
TS-08-296-791-5

100

1

Query Match	Best Local Similarity
3	19

Matches 223; 'Conservative'

10

84 GDHYHYNGKVPY-----
QY 111111

236 GDAYTYGIAGTPYKVNHE

100

120 NEVKGGYIIKVDGKYVYV

Query Match	3.38;	Score 177;	DB 4;	Length 1702;
Best Local Similarity	19.68;	Pred. No. 0.00083;		
Matches 223;	Conservative 141;	Mismatches 431;	Indels 342;	Gaps 53

[illegible]

Db	505	TVLNDKQVDPNSTYGFHGGRLDANGNLT--TFEHRINDG-----ARL 549
Oy	394	VRHGDHEHTYKSNQIQOPLPNNSLAT-PSPLPINPQTSHE-----KHEEDG-----441
Db	550	VNH-----NISKSTVITIGESLITDPNITLTPYNIADPEDNRYAARRIKDGQUL 601
Oy	442	-----YGDANRLIEDSGSVMSGHONHFEKKDLJEEQIKR-AQKHLEVTSHNGLD 496
Db	602	NLENTYIALKKGASTSELPKNSGSENEWMLYMKTSDAKRANVNHINBERM-NGBN 659
Oy	497	SLSSHEQ-----DYPNAKEMDL-----DKRIE-----KXA 524
Db	660	GYFDEEGKNNGNLWYFKEKSDQNRLLGNGLNGDLKVEKGLFLSGRP7PHABDIA 719
Oy	525	GIMK-----QYGVARESTYV-----NKEKALITPHGDHHAAD---PLDEK 563
Db	720	GISSTKDHQFAENNEVVDWDIRNFRKATNINVMNLTLSGRVNAITSNITASDA 779
Oy	564	PVGGHSHSNTELFKPEGVAKKEGNV-----YTGEELTNVYNLLKNTFNNQNTLAN 618
Db	780	KVHIGY-----KAGDVYCRSDYTG-YTCTTDKLSDAALNSPATNVS 822
Oy	619	GQKRVSSFSPEPELEKKIGIMLVKLTTPDG---KYLERVSGVGEVGNIANEELDQ 674
Db	823	GNVMS-----GNAFVYLGKLNLFGTISGNSQSVRLTSHMLNGDSVNOJLNDKGII 878
Oy	675	YLPQOTRKTYIASD---YEVYSTGTFYPTSLATKMASQITLTFPHAGDYYLKVNOF 731
Db	879	HLNQNQDANRWTYNTLTVNSLGSNSFYLLDLSKNDQKVVYKSAAGNPTLQVART 938
Oy	732	AVPRGDTALVAFEDFEGHNYLENNRYKVEIKLPJRLNGQTTFRAGNRIPYFMANAYL 791
Db	939	GEP---TKNELTLPDA--SNA--TRNNLNVSLY-----GNVYDLGAKRYLR 978
Oy	792	DNOSTYVEVPILEKENQ---DKPST-----LP 817
Db	979	NVNGRYDLYNPEVEKRNQOVDJTNTITPNNIQADVSVSNNEELARVETPVPPAPATP 1038
Oy	818	QFRNKAKOENSKIDEX-----VEEKTSER--VEEKESJENGNTS 856
Db	1039	SETTEVAENSQOSKSVKEKNEDATETTAONGEVAEKEKPVKANTQNTNEVAQSGSETE 1098
Oy	857	NSLLEVPYTPVPO-ERKAFASYSYMKLENYLFNDGQIEIYLPSEGYIKKNMADFGCE 915
Db	1099	ETQOTTEIKERANKYKEKAKAVEKEEKAYTER-----DEQOE 1134
Oy	916	APQNGENKP-----SENGKYSTGTVENQ-----TENKPADSL-- 949
Db	1135	APQMASTSPKQAPKAPKEVSTDTKYEEOYVQAOPQOSTYVAALAEATSPNSKPALEOP 1194
Oy	950	PEAPNEKVPV---ENSTQGNLNPNGNVSQDMDPLALEAPV---DPQOELE 999
Db	1195	SEKTNABPVPVSKNOTETATDQPEREKTAIVETEKTOEPQVYASQASPDQOSE 1251

RESULT 15
PCT-US95-10661A-5
Sequence 5. Application PC/TUS9510661A
GENERAL INFORMATION:
APPLICANT: Washington University, et al.
TITLE OF INVENTION: Hemophilus Adherence and Penetration Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Abritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/10661A
 FILING DATE: 16-AUG-1995
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/296,791
 FILING DATE: 25-AUG-1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Treacartin, Richard F.
 REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: FP-59941/RFT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELE: 910 2772299
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1702 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 PCT-US95-10661A-5

Query Match 3.38; Score 177; DB 5; Length 1702;
 Best Local Similarity 19.68; Pred. No. 0.00083;
 Matches 223; Conservative 141; Mismatches 431; Indels 342; Gaps 53;

84 GDHNYNGKVPY-----DALF-----SE-----LMDP--NTQ-LKAD-----IV 119
 236 GDAITGAGIPYKVNHNENGLIFGSKSEHSPKGLISDPLTNVAVLGDSGSPLEFY 295
 120 NEVKGGLIRKVGKYYLLDAHADNVRTKDELINROKOEHVK--DNEKYNNAVAAVS 176
 296 DREKGMFL--GSY-----DFAGYNKKSMQENNIYKPEFAKTVLDKDTAGSLIG---- 344
 177 QGRITNDGYFPNPADILEDNAYIYPHGHYHYIRKSDLSASELAARKHLAKKNOP 236
 345 -----SNTQYMNPP-----TGKTSVLSNGSESLNVLFD--SSODTDSKKNHKKSV-- 389
 237 SOLYSSTASDN-----TQSVAKGSTKPRANKSENLOSLKELYDPS 280
 390 -TLNGSGTLTANNIDGAGGLFEGDEYKGTSDSTTKAGVSVADGKTYTKVHNPK 448
 281 AQRYS-SDGLVFPAPAIISRTPNGVAIPGDNHYHFIYSKLSALEKIAKVPISGTS 339
 449 SDRLAKIGKTLI-----VEGGENKGLKVGDTVILKQADANNKYKAFSGVIGSRS 504
 340 TVSTN---AKPNEVSSL--GSLSSNPSSLTSEKLSASDGTTFNPKDIVEETATAYI 393
 505 TVVLNDKQVDPNSIYFGFRGRLDANGNNL-TFEHIRNIDG-----ARL 549
 394 VRHGDHFIYKSNQIQPLPNSLAT--PSPSLPDPGTSHE-----KHEEDG----- 441
 550 VNH-----NTSKTSTVTITGSLTTPDTPTIPYINADPEDNPYAFRIKIDGGOLYL 601
 442 ----YGFANRIIAEDSGFVMSGHDNHYFFKKDLTEEOIKA--AKHLEEVKTSNGLD 496
 602 NLENTTYAALFKGASTSELPKNGSESNENLTKGTSDAKRVNMHINHERM--NGFN 659
 497 SLSSHEQ-----DYPGNAKEMKD-----DKTIE-----KIA 524
 660 GYFGEDEKNGNINLVTFKGESEONRFLTGTNMLNGDLKVEKGTLLFLSGRPTPHARDIA 719
 525 GIMK-----QYGVKRESIYV-----NKEKNAIYPHGDHHDH--PIDBK 563
 720 GISSTKQDQHAENNEVVEDMIRNFKATNINVTNNATLISGRNVAANTSNITASDNA 779
 564 PVGIGHSNTELFKPEEGVAKKGNKY-----YTGEELTNVVMILKNSFPNNQFTLAN 618
 780 KVHIGY-----KAGDITCVASDITG--YVCTCTDKLSDKALNSFNATNVS 822

619 GQKVSFSPPELEKIGINMLVLLTPDG-----KYLEVSGKVFGBGVNIANFELDOP 674
 823 GNNVLS-----GNNFVIGKANTIGTIGTGNQVRLNENHMLTGDSNNQNLMDGHI 878
 675 YLPGQTFKTYIASD---YEVSTOGFTVPTSLAYMASQTFYPPHAGDTYLRNPPQF 731
 879 HLNAQNDANVTYNTVNTLVNSLSGNGSFYLLDLSNKGGRVVYTKSATGNTFTLQVADKT 938
 732 AVRGDTALVRVDEFGNNALENNYKVGELKLPRLNGOTTRTAGNKIPVTFMANAVYL 791
 939 GEP--TNEELTLEDA--SNA--TRNNLNVSIV-----GNTVDLGMKTKLR 978
 792 DNSTYIVEPPILEKENOT-----DKPSI-----LP 817
 979 NVNKRIVLYNPEVEKRNQYDTTITPNNIQADVPSVPSNNERIANVETPPPPAPATP 1038
 818 QFRKNAQENSKLDER-----VEPPTSEK--VEKELSTGSGTS 856
 1039 SETTEVAENSKQSKVEKNEQDATETTAQNGEVAEAKPSVANTQTNVVAQSGSETE 1098
 857 NSTLEEPTVDPVQ--EKVAFPAESYGKLENVLFNMDGTIELYPSGEVIRKKNADPTGE 915
 1099 ETOTTEIKETAKVEKEKAKVEKEKAKVEK-----DEIOE 1134
 916 APOGNGENKP-----SENGKYSTGVENOP-----TENKPADSL-- 949
 1135 APOMASTSPROAKPAPAKREYSTDKVEETOYQAOPOQOSTVVAALAPSPRSKRAELETOP 1194
 950 PEAPNEKPVPR--ENSTNGMLNPEGNVGSDPMLDPALEAPAV---DVOEKE 999
 1195 SEKTNAPAPVPVASKNQTENTTDOPTEREKTAKEETEKOEPQVVAQASPKROOSE 1251

Search completed: September 26, 2001, 22:31:17
 Job time: 1541 sec